DOI:
10.1016/j.jeconom.2016.11.001

Document Version
Peer reviewed version

Citation for published version (APA):
Massacci, D. (2017). Least squares estimation of large dimensional threshold factor models. *JOURNAL OF ECONOMETRICS*, *197*(1), 101-129. https://doi.org/10.1016/j.jeconom.2016.11.001

Least Squares Estimation of Large Dimensional Threshold Factor Models*

Daniele Massacci

Bank of England

November 13, 2016

Abstract

This paper studies large dimensional factor models with threshold-type regime shifts in the loadings. We estimate the threshold by concentrated least squares, and factors and loadings by principal components. The estimator for the threshold is superconsistent, with convergence rate that depends on the time and cross-sectional dimensions of the panel, and it does not affect the estimator for factors and loadings: this has the same convergence rate as in linear factor models. We propose model selection criteria and a linearity test. Empirical application of the model shows that connectedness in financial variables increases during periods of high economic policy uncertainty.

JEL classification: C12, C13, C33, C52, G10.

Keywords: Large Threshold Factor Model, Least Squares Estimation, Model Selection, Linearity Testing, Connectedness.

*This work was carried out when the author was Franco Modigliani Research Fellow in Economics and Finance at the Einaudi Institute for Economics and Finance (EIEF); it was revised after the author joined the Bank of England. The views in this paper are the author's and do not necessarily reflect those of the Bank of England, or its policy committees. The author really is highly indebted to Marco Lippi for introducing him to factor models and for several enlightening conversations. Kind hospitality by the Department of Finance at Bocconi University is acknowledged. This paper benefits from comments from seminar participants at University of Rome Tor Vergata, Bocconi University and the Bank of Italy; from suggestions from conference participants at the Vienna Workshop on High-Dimensional Time Series in Macroeconomics and Finance, and at the IAAE 2015 Annual Conference; and from conversations with Domenico Giannone, Alessandro Giovannelli, Hashem Pesaran and Stefano Soccorsi. The author thanks Jianqing Fan (the co-editor), the associate editor, and two anonymous referees for the insightful comments. Errors and omissions are the author's own responsibility. Financial support from the Associazione Borsisti Marco Fanno and from UniCredit and Universities Foundation is gratefully acknowledged. Address correspondence to Daniele Massacci, Bank of England, Threadneedle Street, London, EC2R 8AH, United Kingdom. E-mail: dm355@cantab.net.

1 Introduction

Factor models are widely used tools to explain the common variations in large scale macroeconomic and financial data. An extensive literature analyzes factor models under the maintained assumption of constant loadings over the entire sample period: see Connor and Korajczyk (1986, 1988, 1993), Forni *et al.* (2000, 2004, 2015), Forni and Lippi (2001), Bai and Ng (2002), Stock and Watson (2002), and Bai (2003) for seminal contributions on linear factor models. Economic models are however unlikely to have constant parameters over time and factor models with time-dependent loadings are called for. Time-dependence in the loadings may be easily implemented through a change-point mechanism: this may be parameterized as either a structural break or a regime shift driven by the threshold principle, depending on the underlying data generating process.

Structural breaks in the loadings may arise as a consequence of events such as technological or policy changes. Several important contributions deal with large dimensional factor models subject to loadings instabilities. Breitung and Eickmeier (2011) show that ignoring breaks leads to overestimation of the number of factors and develop statistical tests for the null hypothesis of stability in the loadings. Bates *et al.* (2013) study the robustness properties of the principal components estimator of the factors under neglected loadings instability. Chen *et al.* (2014), Han and Inoue (2015) and Yamamoto and Tanaka (2015) develop further statistical tools to detect breaks. Chen (2015) considers least squares estimation of the break date. Cheng *et al.* (2015) propose shrinkage estimation of large dimensional factor models with structural breaks.

Regime shift representations of the dependent variables are suitable when "history repeats", as with financial returns (Timmermann (2008), and Ang and Timmermann (2012)). Ng and Wright (2013) introduce a threshold mechanism in large dimensional factor models to simulate data and investigate the effects of nonlinearities on business cycle dynamics[1]. We take Ng and Wright (2013) intuition as a starting point and propose a large dimensional factor model with regime changes in the loadings governed by the threshold principle. We let the threshold value be unknown and focus on estimation, model selection and linearity testing. To the very best of our knowledge, we are the first to tackle this problem.

Let R^0 be the true number of factors. Under the maintained assumption that R^0 is known, we propose to estimate the threshold value by concentrated least squares, and factors and loadings by

[1] See Ng and Wright (2013), p. 1147.

principal components (Hansen (2000), and Bai and Ng (2002)). We obtain a number of novel theoretical results. Let N and T denote the cross-sectional and time series dimensions, respectively. We first provide sufficient conditions to ensure that our model is identified from a linear factor model: formally, for $0.5 < \alpha^0 \leq 1$, we require that *at least* a fraction $O\left(N^{\alpha^0}\right)$ of the N cross-sectional units experiences a regime shift in the loadings, so that the shift resists to the aggregation induced by the principal components estimator. We then show that the estimator for the threshold parameter is consistent at a rate equal to $N^{\alpha^0}T$: this depends on the time series dimension T and the number of cross-sectional units N^{α^0} subject to the threshold effect. The convergence rate monotonically increases in α^0 and it is such that $\sqrt{N}T < N^{\alpha^0}T \leq NT$: this shows the direct relationship between identification of the model and convergence rate of the estimator for the threshold. As a consequence of this superconsistency property, we finally show that the principal components estimator for both regime-specific loadings and factors have convergence rate equal to $C_{NT} = \min\left\{\sqrt{N}, \sqrt{T}\right\}$: despite the threshold effect, the convergence rate C_{NT} is equal to the one derived in Bai and Ng (2002) for linear factor models.

We next let the true number of factors R^0 be unknown so that it has to be estimated. Breitung and Eickmeier (2011) show that structural instability in the loadings leads to a factor representation with a higher dimensional factor space: due to an analogy argument, the same issue arises when a regime shift drives time variation in the loadings. Since the convergence rate C_{NT} of the estimator for loadings and factors is the same as in linear factor models, we make Bai and Ng (2002) information criteria robust to the threshold effect by accounting for the induced higher dimensional factor space representation.

As a last theoretical contribution, we propose a linearity test. Following Chen *et al.* (2014), and Han and Inoue (2015), we check whether the covariance matrix of the estimated factors is regime-dependent: we use the regression approach of Chen *et al.* (2014) and extend Hansen (1996) seminal contribution to derive the asymptotic distribution of the test statistic under the null hypothesis of linearity.

We finally show how our theoretical framework may be used to measure connectedness in financial markets (Acharya *et al.* (2010), Billio *et al.* (2012), Engle and Kelly (2012), Diebold and Yilmaz (2014), and Adrian and Brunnermeier (2016)). We extend Billio *et al.* (2012) measure based on principal components analysis to allow for regime-specific connectedness. Using Baker *et al.* (2016) index of economic policy uncertainty as threshold variable, we show that connectedness in financial markets increases during periods of high uncertainty: this may be relevant for risk measurement and management.

The remainder of the paper is organized as follows. Section 2 describes the model. Section 3 deals with estimation. Section 4 looks at model selection. Section 5 develops a linearity test. Section 6 performs a Monte Carlo analysis. Section 7 provides an empirical application. Section 8 outlines directions for future research. Finally, Section 9 concludes. Appendix A provides technical proofs.

Concerning notation, $\mathbb{I}(\cdot)$ denotes the indicator function; given a square matrix $\mathbf{A}$, $\mathrm{tr}(\mathbf{A})$ denotes the trace of $\mathbf{A}$; the norm of a generic matrix $\mathbf{A}$ is $\|\mathbf{A}\| = [\mathrm{tr}(\mathbf{A}'\mathbf{A})]^{1/2}$; for a given scalar A, $|A|$, $\mathbf{I}_A$ and $\mathbf{0}_A$ are the absolute value of A, the $A \times A$ identity matrix and the zero matrix, respectively; $\xrightarrow{p}$ denotes convergence in probability; $\xrightarrow{d}$ denotes convergence in distribution; $\Rightarrow$ denotes weak convergence with respect to the uniform metric.

2 The Approximate Threshold Factor Model

We consider the model

$$\mathbf{x}_t = \mathbb{I}(z_t \leq \theta)\,\mathbf{\Lambda}_1 \mathbf{f}_t + \mathbb{I}(z_t > \theta)\,\mathbf{\Lambda}_2 \mathbf{f}_t + \mathbf{e}_t, \quad t = 1, \ldots, T, \tag{1}$$

where T denotes the time series dimension of the available sample; $\mathbf{x}_t = (x_{1t}, \ldots, x_{Nt})' \in \Re^N$ is the $N \times 1$ vector of observable dependent variables; $\mathbf{f}_t = (f_{1t}, \ldots, f_{Rt})' \in \Re^R$ is the $R \times 1$ vector of latent factors; $z_t \in \Re$ is an observable covariate and θ is the unknown threshold value; $\mathbf{e}_t = (e_{1t}, \ldots, e_{Nt})' \in \Re^N$ is the $N \times 1$ vector of idiosyncratic errors; $\mathbf{\Lambda}_j = (\boldsymbol{\lambda}_{j1}, \ldots, \boldsymbol{\lambda}_{jN})'$ is the $N \times R$ matrix of factor loadings with $i-th$ row defined as $\boldsymbol{\lambda}_{ji} = (\lambda_{ji1}, \ldots, \lambda_{jiR})'$, for $j = 1, 2$ and $i = 1, \ldots, N$.

The model in (1) belongs to the class of threshold models proposed in Tong and Lim (1980): see Tsay (1989, 1998), Chan (1993) and Hansen (1996, 1999, 2000) for methodological contributions; and Hansen (2011) for a survey of the literature. According to the threshold principle introduced in Pearson (1900), the regime prevailing at time t depends on the position of z_t with respect to the unknown threshold θ. Ng and Wright (2013) simulate data from a large dimensional threshold factor model to investigate the effects of nonlinearities on business cycle dynamics[2]: we explicitly focus on estimation, model selection and linearity testing. Our results extend to the case in which the threshold variable is more generally defined as a linear combination of covariates (Massacci (2014)): this would be relevant when the driver

[2] See Ng and Wright (2013), p. 1147.

of the regimes is not *a priori* known.

The model in (1) extends large dimensional linear factor models to allow for a threshold effect on the loadings. Given Assumption C3 stated in Section 3.1 below, we follow Chamberlain and Rothschild (1983) and allow for some degree of correlation in the idiosyncratic components within each regime: (1) then is an *approximate threshold factor model*; it is more general than an *exact threshold factor model*, which would extend the arbitrage pricing theory of Ross (1976) and would not allow for any correlation in the idiosyncratic components in any regime.

3 Estimation

As in Stock and Watson (2002), we study estimation of (1) under the assumption that the true number of factors R^0 (i.e., the true dimension of $\mathbf{f}_t$) is known. We extend the theory in Bai and Ng (2002) based on principal components estimation to allow for concentrated least squares estimation, as motivated in Hansen (2000) for threshold regressions. The plan is as follows: Section 3.1 states the assumptions; Section 3.2 deals with identification; Section 3.3 describes the principal components estimator; Section 3.4 proves the consistency of the estimator; and Section 3.5 derives the convergence rates.

3.1 Assumptions

We group the assumptions into three sets, depending on the role they play to identify and estimate the model, and to derive the convergence rates. Let $\mathbb{I}_{1t}(\theta) = \mathbb{I}(z_t \leq \theta)$ and $\mathbb{I}_{2t}(\theta) = \mathbb{I}(z_t > \theta)$. For $j = 1, 2$, denote $\mathbf{\Lambda}_j^0 = \left(\boldsymbol{\lambda}_{j1}^0, \ldots, \boldsymbol{\lambda}_{jN}^0\right)'$, θ^0 and $\mathbf{f}_t^0$ the true values of $\mathbf{\Lambda}_j$, θ and $\mathbf{f}_t$, respectively. Define $\mathbf{f}_{jt}^0(\theta) = \mathbb{I}_{jt}(\theta)\,\mathbf{f}_t^0$, for $j = 1, 2$ and $t = 1, \ldots, T$, and let $\boldsymbol{\delta}_i^0 = \boldsymbol{\lambda}_{2i}^0 - \boldsymbol{\lambda}_{1i}^0$, for $i = 1, \ldots, N$.

3.1.1 Identification

Assumption I - Threshold Factor Model. For $0.5 < \alpha^0 \leq 1$, $\boldsymbol{\delta}_i^0 \neq \mathbf{0}$ for $i = 1, \ldots, N^{\alpha^0}$, and $\sum_{i=N^{\alpha^0}+1}^{N} \boldsymbol{\delta}_i^0 = O(1)$.

Assumption I requires that *at least* a fraction $O\left(N^{\alpha^0}\right)$ of the N series experiences a threshold effect, for $0.5 < \alpha^0 \leq 1$: this follows up on Bates *et al.* (2013), who show that if at most $O\left(N^{0.5}\right)$ series undergo a break then the principal components estimator as applied to the misspecified linear model achieves the same Bai and Ng (2002) convergence rate. Assumption I ensures that enough series experience a regime shift so that (1) is identified from a linear factor model when factors and loadings are estimated by principal components. As shown in Theorem 3.4 below, α^0 affects the convergence rate of the estimator for θ^0: the higher the former, the faster the latter. In this paper we do not aim at estimating α^0 and leave this interesting issue to future research.

3.1.2 Consistency

Assumption C1 - Factors. $\mathrm{E}\left\|\mathbf{f}_t^0\right\|^4 < \infty$; for $j = 1,2$, $T^{-1}\sum_{t=1}^T \mathbf{f}_{jt}^0(\theta)\,\mathbf{f}_{jt}^0\left(\theta^0\right)' \xrightarrow{p} \mathbf{\Sigma}_{j\mathbf{f}}^0\left(\theta,\theta^0\right)$ as $T\to\infty$ for all θ and some positive definite matrix $\mathbf{\Sigma}_{j\mathbf{f}}^0\left(\theta,\theta^0\right)$.

Assumption C2 - Factor Loadings. For $j = 1,2$ and $i = 1,\ldots,N$, $\left\|\boldsymbol{\lambda}_{ji}^0\right\| \le \bar{\lambda} < \infty$, and $\left\|\mathbf{\Lambda}_j^{0\prime}\mathbf{\Lambda}_j^0/N - \mathbf{D}_{\mathbf{\Lambda}_j}^0\right\| \to 0$ as $N\to\infty$ for some $R^0\times R^0$ positive definite matrix $\mathbf{D}_{\mathbf{\Lambda}_j}^0$.

Assumption C3 - Time and Cross-Section Dependence and Heteroskedasticity. There exists a positive $M<\infty$ such that for $j=1,2$, for all θ and for all (N,T),

(a) $\mathrm{E}\left(e_{it}\right) = 0$ and $\mathrm{E}\left|e_{it}\right|^8 \le M$;

(b) $\mathrm{E}\left[\mathbb{I}_{jt}(\theta)\,\mathbb{I}_{jv}(\theta)\,e_{it}e_{iv}\right] = \tau_{jitv}(\theta)$ with $\left|\tau_{jitv}(\theta)\right| \le \left|\tau_{jtv}\right|$ for some τ_{jtv} and for all i, and $T^{-1}\sum_{t=1}^T\sum_{v=1}^T\left|\tau_{jtv}\right| \le M$;

(c) $\mathrm{E}\left[T^{-1}\sum_{t=1}^T \mathbb{I}_{jt}(\theta)\,e_{it}e_{lt}\right] = \sigma_{jil}(\theta)$, $\left|\sigma_{jll}(\theta)\right| \le M$ for all l, and $N^{-1}\sum_{i=1}^N\sum_{l=1}^N\left|\sigma_{jil}(\theta)\right| \le M$;

(d) $\mathrm{E}\left|T^{-1/2}\sum_{t=1}^T \mathbb{I}_{jt}(\theta)\,e_{it}e_{lt} - \mathrm{E}\left[\mathbb{I}_{jt}(\theta)\,e_{it}e_{lt}\right]\right|^4 \le M$ for every (i,l).

Assumption C4 - Weak Dependence between $\mathbf{f}_t^0$, z_t and e_{it}. There exists some positive constant $M<\infty$ such that for all θ and for all (N,T),

$$\mathrm{E}\left\{N^{-1}\sum_{i=1}^N\left\|T^{-1/2}\left[\sum_{t=1}^T \mathbb{I}_{jt}(\theta)\,\mathbf{f}_t^0 e_{it}\right]\right\|^2\right\} \le M, \quad j=1,2.$$

Assumptions C1 to C4 are the natural extensions of Assumptions A to D imposed on linear factor models in Bai and Ng (2002) and accommodate the threshold effect. Assumption C1 restricts the sequences $\left\{\mathbf{f}_t^0\right\}_{t=1}^T$ and $\left\{z_t\right\}_{t=1}^T$ so that appropriate second moments exist; it also imposes a full rank condition that excludes multicollinearity in the factors. According to Assumption C2, factor loadings are nonstochastic and each factor has a nonnegligible effect on the variance of $\mathbf{x}_t$ within each regime.

Under Assumption C3, limited degrees of time-series and cross-section dependence in the idiosyncratic components as well as heteroskedasticity are allowed. Finally, Assumption C4 provides an upper bound to the degree of dependence between the factors, z_t and the idiosyncratic components: Assumption C4 is stronger than Assumption D in Bai and Ng (2002), which only bounds the dependence between the factors and the idiosyncratic components. Although we deal with a panel structure, we do not require the threshold variable z_t to be strictly exogenous as in Assumption 2 in Hansen (1999): in particular, z_t is allowed to be predetermined and equal to some lagged value of one of the elements of $\mathbf{x}_t$.

3.1.3 Convergence Rates

Define $\mathbf{D}_{\mathbf{f}}^0(\theta) = \mathrm{E}\left(\mathbf{f}_t^0\mathbf{f}_t^{0\prime}\,|z_t = \theta\right)$ and denote by $f_{\mathcal{Z}}(z_t)$ the density function of z_t.

Assumption CR - Stationarity, Moment Bound, Continuity and Full Rank.

- **(a)** $\left\{\mathbf{f}_t^0, z_t, \mathbf{e}_t\right\}_{t=1}^T$ is strictly stationary, ergodic and $\rho-$mixing, with $\rho-$mixing coefficients satisfying $\sum_{m=1}^{\infty}\rho_m^{1/2} < \infty$;
- **(b)** For all θ, $\mathrm{E}\left(\left\|\mathbf{f}_t^0 e_{it}\right\|^4|z_t = \theta\right) \leq C$ and $\mathrm{E}\left(\left\|\mathbf{f}_t^0\right\|^4|z_t = \theta\right) \leq C$ for some $C < \infty$ and for $i = 1, \ldots, N$, and $f_{\mathcal{Z}}(\theta) \leq \bar{f} < \infty$;
- **(c)** $f_{\mathcal{Z}}(\theta)$ and $\mathbf{D}_{\mathbf{f}}^0(\theta)$ are continuous at $\theta = \theta^0$;
- **(d)** $\boldsymbol{\delta}_i^{0\prime}\mathbf{D}_{\mathbf{f}}^0\left(\theta^0\right)\boldsymbol{\delta}_i^0 > 0$, $i = 1, \ldots, N^{\alpha^0}$ and $0.5 < \alpha^0 \leq 1$, and $\sum_{i=N^{\alpha^0}+1}^{N}\boldsymbol{\delta}_i^{0\prime}\mathbf{D}_{\mathbf{f}}^0\left(\theta^0\right)\boldsymbol{\delta}_i^0 = O(1)$; $f_{\mathcal{Z}}(\theta) > 0$ for all θ.

Assumption CR is analogous to Assumption 1 in Hansen (2000). Assumption CR(a) restricts the memory of the sequence $\left\{\mathbf{f}_t^0, z_t, \mathbf{e}_t\right\}_{t=1}^T$; it excludes trends and integrated processes. Assumption CR(b) gives conditional moment bounds. Assumption CR(c) imposes a continuous support on z_t. The full-rank condition in Assumption CR(d) strengthens Assumption I and rules out the "continuous threshold" set up of Chan and Tsay (1998), which arises in the one-factor model when the scalar factor f_t^0 equals the threshold variable z_t and $\theta^0 = 0$: in this case, $\delta_i^0\mathrm{E}\left(f_t^0 f_t^0\,|z_t = \theta^0\right)\delta_i^0 = \delta_i^0\mathrm{E}\left(f_t^0 f_t^0\,|f_t^0 = \theta^0\right)\delta_i^0 = 0$, for $i = 1, \ldots, N$, and Assumption CR(d) is violated.

3.2 Identification

Let $\mathbf{\Delta}^0 = \left(\boldsymbol{\delta}_1^0, \ldots, \boldsymbol{\delta}_N^0\right)'$ and write the data generating process of $\mathbf{x}_t$ as $\mathbf{x}_t = \mathbf{\Lambda}_1^0 \mathbf{f}_t^0 + \mathbb{I}_{2t}\left(\theta^0\right) \mathbf{\Delta}^0 \mathbf{f}_t^0 + \mathbf{e}_t$. Define $\mathbf{F}^0 = \left(\mathbf{f}_1^0, \ldots, \mathbf{f}_T^0\right)$ and denote $\tilde{\mathbf{\Lambda}}_1 = \left(\tilde{\boldsymbol{\lambda}}_{11}, \ldots, \tilde{\boldsymbol{\lambda}}_{1N}\right)'$ the principal components estimator for $\mathbf{\Lambda}_1^0$ from the misspecified linear factor model $\mathbf{x}_t = \mathbf{\Lambda}_1 \mathbf{f}_t + \mathbf{e}_t$. Let $\tilde{\mathbf{V}}_1$ be the $R^0 \times R^0$ diagonal matrix of the first R^0 largest eigenvalues of $\hat{\mathbf{\Sigma}}_{\mathbf{x}} = (NT)^{-1} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'$ in decreasing order: the underlying optimization problem requires the normalization $N^{-1} \tilde{\mathbf{\Lambda}}_1' \tilde{\mathbf{\Lambda}}_1 = \mathbf{I}_{R^0}$. The following theorem states the properties of $\tilde{\mathbf{\Lambda}}_1$.

Theorem 3.1 *There exists a $R^0 \times R^0$ rotation matrix $\tilde{\mathbf{H}}_1$ with* $\operatorname{rank}\left(\tilde{\mathbf{H}}_1\right) = R^0$ *such that*

$$B_{NT}^2 \left(\frac{1}{N} \sum_{i=1}^N \left\| \tilde{\boldsymbol{\lambda}}_{1i} - \tilde{\mathbf{H}}_1' \boldsymbol{\lambda}_{1i}^0 \right\|^2 \right) = O_p(1),$$

as $N, T \to \infty$, where

$$B_{NT} = \min \left\{ \sqrt{N}, \sqrt{T}, N^{1-\alpha^0} \right\}$$

and

$$\tilde{\mathbf{H}}_1 = \frac{\mathbf{F}^0 \mathbf{F}^{0\prime}}{T} \frac{\mathbf{\Lambda}_1^{0\prime} \tilde{\mathbf{\Lambda}}_1}{N} \tilde{\mathbf{V}}_1^{-1}.$$

Theorem 3.1 shows that the average squared deviations between the loadings estimated using a linear factor model and those that lie in the true loading space vanish as $N, T \to \infty$ at a rate equal to B_{NT}^2, which drives identification. Under Assumption I, the model in (1) is identified from the linear factor model as the rate of convergence $N^{1-\alpha^0}$ of the principal components estimator is slower than it would be under correct linear model specification: the model in (1) would not be identified from a linear factor model if $0 \leq \alpha^0 \leq 0.5$, since in this case $B_{NT}^2 = \min\{N, T\}$, as derived in Bai and Ng (2002). If $\alpha^0 = 1$ and all cross-sectional units are subject to threshold effect, $B_{NT}^2 = 1$ and the principal components estimator from the misspecified linear model is asymptotically biased. As proved in Theorem 3.4, the parameter α^0 regulates the convergence rate of the estimator for the unknown threshold value θ^0: this result shows the connection between identification strength and estimation precision.

3.3 Principal Components Estimation

We estimate factors and loadings by principal components, and θ^0 by concentrated least squares: see Bai and Ng (2002) and Hansen (2000), respectively. Define the $N \times 2R^0$ matrix of loadings $\mathbf{\Lambda} = (\mathbf{\Lambda}_1, \mathbf{\Lambda}_2)$

and the $R^0 \times T$ matrix of factors $\mathbf{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_T)$. Let $\mathbf{\Lambda}^0 = \left(\mathbf{\Lambda}_1^0, \mathbf{\Lambda}_2^0\right)$ be the true value of $\mathbf{\Lambda}$. The objective function in terms of $\mathbf{\Lambda}$, $\mathbf{F}$ and θ is the sum of squared residuals (divided by NT)

$$S(\mathbf{\Lambda}, \mathbf{F}, \theta) = (NT)^{-1} \sum_{t=1}^{T} \left[\mathbf{x}_t - \mathbb{I}_{1t}(\theta)\mathbf{\Lambda}_1\mathbf{f}_t - \mathbb{I}_{2t}(\theta)\mathbf{\Lambda}_2\mathbf{f}_t\right]' \left[\mathbf{x}_t - \mathbb{I}_{1t}(\theta)\mathbf{\Lambda}_1\mathbf{f}_t - \mathbb{I}_{2t}(\theta)\mathbf{\Lambda}_2\mathbf{f}_t\right] : \tag{2}$$

the estimators $\hat{\mathbf{\Lambda}} = \left(\hat{\mathbf{\Lambda}}_1, \hat{\mathbf{\Lambda}}_2\right)$, $\hat{\mathbf{F}} = \left(\hat{\mathbf{f}}_1, \ldots, \hat{\mathbf{f}}_T\right)$ and $\hat{\theta}$ for $\mathbf{\Lambda}^0$, $\mathbf{F}^0$and θ^0, respectively, with $\hat{\mathbf{\Lambda}}_j = \left(\hat{\boldsymbol{\lambda}}_{j1}, \ldots, \hat{\boldsymbol{\lambda}}_{jN}\right)'$, for $j = 1, 2$, jointly solve

$$\hat{\mathbf{\Lambda}}, \hat{\mathbf{F}}, \hat{\theta} = \arg\min_{\mathbf{\Lambda}, \mathbf{F}, \theta} S(\mathbf{\Lambda}, \mathbf{F}, \theta).$$

For given $\mathbf{\Lambda}$ and θ, and subject to $N^{-1}\left(\mathbf{\Lambda}_j'\mathbf{\Lambda}_j\right) = \mathbf{I}_{R^0}$, for $j = 1, 2$, from (2) we have

$$\hat{\mathbf{f}}_t(\mathbf{\Lambda}, \theta) = N^{-1}\left[\mathbb{I}_{1t}(\theta)\mathbf{\Lambda}_1 + \mathbb{I}_{2t}(\theta)\mathbf{\Lambda}_2\right]' \mathbf{x}_t, \quad t = 1, \ldots, T : \tag{3}$$

replacing $\mathbf{f}_t$ in (2) with $\hat{\mathbf{f}}_t(\mathbf{\Lambda}, \theta)$ obtained in (3) leads to the concentrated objective function

$$S_{\mathbf{F}}(\mathbf{\Lambda}, \theta) = (NT)^{-1} \sum_{t=1}^{T} \mathbf{x}_t' \left\{\mathbf{I}_N - N^{-1}\left[\mathbb{I}_{1t}(\theta)\mathbf{\Lambda}_1\mathbf{\Lambda}_1' + \mathbb{I}_{2t}(\theta)\mathbf{\Lambda}_2\mathbf{\Lambda}_2'\right]\right\} \mathbf{x}_t, \tag{4}$$

and the estimators for $\mathbf{\Lambda}^0$ and θ^0 jointly solve

$$\hat{\mathbf{\Lambda}}, \hat{\theta} = \arg\min_{\mathbf{\Lambda}, \theta} S_{\mathbf{F}}(\mathbf{\Lambda}, \theta).$$

From (4), the estimator for $\mathbf{\Lambda}^0$ for given θ is defined as

$$\hat{\mathbf{\Lambda}}(\theta) = \left[\hat{\mathbf{\Lambda}}_1(\theta), \hat{\mathbf{\Lambda}}_2(\theta)\right] = \arg\max_{\mathbf{\Lambda}} V_{\mathbf{F}}(\mathbf{\Lambda}, \theta), \tag{5}$$

where

$$\begin{aligned} V_{\mathbf{F}}(\mathbf{\Lambda}, \theta) &= (NT)^{-1} \sum_{t=1}^{T} \left\{\mathbf{x}_t' \left[\mathbb{I}_{1t}(\theta)(\mathbf{\Lambda}_1\mathbf{\Lambda}_1') + \mathbb{I}_{2t}(\theta)(\mathbf{\Lambda}_2\mathbf{\Lambda}_2')\right]\mathbf{x}_t\right\} \\ &= (NT)^{-1} \left\{\operatorname{tr}\left\{\mathbf{\Lambda}_1' \left[\sum_{t=1}^{T} \mathbb{I}_{1t}(\theta)\mathbf{x}_t\mathbf{x}_t'\right]\mathbf{\Lambda}_1\right\} + \operatorname{tr}\left\{\mathbf{\Lambda}_2' \left[\sum_{t=1}^{T} \mathbb{I}_{2t}(\theta)\mathbf{x}_t\mathbf{x}_t'\right]\mathbf{\Lambda}_2\right\}\right\}. \end{aligned}$$

The problem

$$\max_{\mathbf{\Lambda}} V_{\mathbf{F}}\left(\mathbf{\Lambda}, \theta\right)$$

is equivalent to

$$\max_{\mathbf{\Lambda}} \left[\mathbf{\Lambda}_1' \hat{\mathbf{\Sigma}}_{1\mathbf{x}}\left(\theta\right) \mathbf{\Lambda}_1 + \mathbf{\Lambda}_2' \hat{\mathbf{\Sigma}}_{2\mathbf{x}}\left(\theta\right) \mathbf{\Lambda}_2\right], \tag{6}$$

where

$$\hat{\mathbf{\Sigma}}_{j\mathbf{x}}\left(\theta\right) = \left[\left(NT\right)^{-1} \sum_{t=1}^{T} \mathbb{I}_{jt}\left(\theta\right) \mathbf{x}_t \mathbf{x}_t'\right], \quad j = 1, 2: \tag{7}$$

for $j = 1, 2$, and for given θ, the estimator for $\mathbf{\Lambda}_j^0$ solving the problem in (6) is $\hat{\mathbf{\Lambda}}_j\left(\theta\right)$, where $\hat{\mathbf{\Lambda}}_j\left(\theta\right)$ is equal to $\sqrt{N}$ times the $N \times R^0$ matrix of eigenvectors of $\hat{\mathbf{\Sigma}}_{j\mathbf{x}}\left(\theta\right)$ corresponding to its largest R^0 eigenvalues. Replacing $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ in (4) with $\hat{\mathbf{\Lambda}}_1\left(\theta\right)$ and $\hat{\mathbf{\Lambda}}_2\left(\theta\right)$ leads to the concentrated sum of squared residuals (divided by NT)

$$S_{\mathbf{F\Lambda}}\left(\theta\right) = \left(NT\right)^{-1} \sum_{t=1}^{T} \mathbf{x}_t' \left\{\mathbf{I}_N - N^{-1}\left[\mathbb{I}_{1t}\left(\theta\right) \hat{\mathbf{\Lambda}}_1\left(\theta\right) \hat{\mathbf{\Lambda}}_1\left(\theta\right)' + \mathbb{I}_{2t}\left(\theta\right) \hat{\mathbf{\Lambda}}_2\left(\theta\right) \hat{\mathbf{\Lambda}}_2\left(\theta\right)'\right]\right\} \mathbf{x}_t: \tag{8}$$

the estimator for θ^0 then solves

$$\hat{\theta} = \arg\min_{\theta} S_{\mathbf{F\Lambda}}\left(\theta\right).$$

Given $\hat{\theta}$, the estimator for $\mathbf{\Lambda}_j^0$ is $\hat{\mathbf{\Lambda}}_j = \hat{\mathbf{\Lambda}}_j\left(\hat{\theta}\right)$, for $j = 1, 2$. Finally, given $\hat{\theta}$ and $\hat{\mathbf{\Lambda}} = \left(\hat{\mathbf{\Lambda}}_1, \hat{\mathbf{\Lambda}}_2\right)$, from (3)

$$\hat{\mathbf{f}}_t = \hat{\mathbf{f}}_t\left(\hat{\mathbf{\Lambda}}, \hat{\theta}\right) = N^{-1}\left[\mathbb{I}_{1t}\left(\hat{\theta}\right) \hat{\mathbf{\Lambda}}_1 + \mathbb{I}_{2t}\left(\hat{\theta}\right) \hat{\mathbf{\Lambda}}_2\right]' \mathbf{x}_t, \quad t = 1, \ldots, T.$$

3.4 Consistency

From Theorem 3.1, the two regimes described in (1) are separately identified under Assumption 1. Define the $R^0 \times T$ matrices of regime-specific factors $\mathbf{F}_j^0\left(\theta\right) = \left[\mathbf{f}_{j1}^0\left(\theta\right), \ldots, \mathbf{f}_{jT}^0\left(\theta\right)\right]$, for $j = 1, 2$, such that $\mathbf{F}_1^0\left(\theta\right) + \mathbf{F}_2^0\left(\theta\right) = \left(\mathbf{f}_1^0, \ldots, \mathbf{f}_T^0\right) = \mathbf{F}^0$ for any θ, and $\mathbf{F}_1^0\left(\theta^0\right) \mathbf{F}_2^0\left(\theta^0\right)' = \mathbf{0}_{R^0}$. Let $\hat{\mathbf{H}}_{jj}\left(\theta\right)$ and $\hat{\mathbf{H}}_{mj}\left(\theta\right)$ be the rotation matrices

$$\hat{\mathbf{H}}_{jj}\left(\theta\right) = \frac{\mathbf{F}_j^0\left(\theta^0\right) \mathbf{F}_j^0\left(\theta\right)'}{T} \frac{\mathbf{\Lambda}_j^{0\prime} \hat{\mathbf{\Lambda}}_j\left(\theta\right)}{N} \hat{\mathbf{V}}_j\left(\theta\right)^{-1}, \quad j = 1, 2, \tag{9}$$

$$\hat{\mathbf{H}}_{mj}\left(\theta\right) = \frac{\mathbf{F}_m^0\left(\theta^0\right) \mathbf{F}_j^0\left(\theta\right)'}{T} \frac{\mathbf{\Lambda}_m^{0\prime} \hat{\mathbf{\Lambda}}_j\left(\theta\right)}{N} \hat{\mathbf{V}}_j\left(\theta\right)^{-1}, \quad j, m = 1, 2, \quad j \neq m, \tag{10}$$

where $\hat{\mathbf{V}}_j(\theta)$ is the $R^0 \times R^0$ diagonal matrix of the first R^0 largest eigenvalues of $\hat{\mathbf{\Sigma}}_{j\mathbf{x}}(\theta)$ defined in (7) in decreasing order: for $\theta = \theta^0$ notice that $\hat{\mathbf{H}}_{jj}(\theta)$ and $\hat{\mathbf{H}}_{mj}(\theta)$ reduce to

$$\hat{\mathbf{H}}_{jj}\left(\theta^0\right) = \frac{\mathbf{F}_j^0\left(\theta^0\right)\mathbf{F}_j^0\left(\theta^0\right)'}{T}\frac{\mathbf{\Lambda}_j^{0\prime}\hat{\mathbf{\Lambda}}_j\left(\theta^0\right)}{N}\hat{\mathbf{V}}_j\left(\theta^0\right)^{-1}, \quad \hat{\mathbf{H}}_{mj}\left(\theta^0\right) = \mathbf{0}_{R^0} \quad j,m = 1,2, \quad j \neq m,$$

and $\hat{\mathbf{H}}_{jj}\left(\theta^0\right)$ becomes a regime-specific rotation matrix analogous to the one derived in Bai and Ng (2002) for linear factor models[3]. The following theorem shows the bias of the principal components estimator induced by the presence of regimes when $\theta \neq \theta^0$.

Theorem 3.2 *There exist $R^0 \times R^0$ matrices $\hat{\mathbf{H}}_{jj}(\theta)$ and $\hat{\mathbf{H}}_{mj}(\theta)$ as defined in* (9) *and* (10), *respectively, with* $\text{rank}\left[\hat{\mathbf{H}}_{jj}(\theta)\right] = R^0$ *for all θ, and* $\text{rank}\left[\hat{\mathbf{H}}_{mj}(\theta)\right] = R^0$ *for $\theta \neq \theta^0$, and $C_{NT} = \min\left\{\sqrt{N}, \sqrt{T}\right\}$, such that*

$$C_{NT}^2\left[\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\boldsymbol{\lambda}}_{ji}(\theta) - \hat{\mathbf{H}}_{jj}(\theta)'\boldsymbol{\lambda}_{ji}^0 - \hat{\mathbf{H}}_{mj}(\theta)'\boldsymbol{\lambda}_{mi}^0\right\|^2\right] = O_p(1), \quad \forall\theta, \quad j,m = 1,2, \quad j \neq m.$$

Theorem 3.2 shows that the presence of regimes adds the asymptotic bias $\hat{\mathbf{H}}_{mj}(\theta)'\boldsymbol{\lambda}_{mi}^0$ to the principal components estimator $\hat{\boldsymbol{\lambda}}_{ji}(\theta)$ for the space $\hat{\mathbf{H}}_{jj}(\theta)'\boldsymbol{\lambda}_{ji}^0$ spanned by $\boldsymbol{\lambda}_{ji}^0$. As in linear factor models, the rate of convergence is equal to $C_{NT}^2 = \min\{N,T\}$ and therefore depends on the panel structure. Taking into account (10), it follows that for $\theta = \theta^0$,

$$C_{NT}^2\left[\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\boldsymbol{\lambda}}_{ji}\left(\theta^0\right) - \hat{\mathbf{H}}_{jj}\left(\theta^0\right)'\boldsymbol{\lambda}_{ji}^0\right\|^2\right] = O_p(1), \quad j = 1,2, \tag{11}$$

which extends the result in Theorem 1 in Bai and Ng (2002) to accommodate the presence of regimes when the threshold θ^0 is known.

Theorem 3.2 plays a key role in proving the following theorem, which states the consistency of $\hat{\theta}$ as an estimator for θ^0.

Theorem 3.3 *Under Assumptions I and C1-C4, $\hat{\theta} \xrightarrow{p} \theta^0$ as $N, T \to \infty$.*

Theorems 3.2 and 3.3 imply a number of results analogous to those collected in Theorem 1 in Stock and Watson (2002): these are stated in Corollary 3.1 below.

[3]See Bai and Ng (2002), p. 213.

Corollary 3.1 *For $j = 1, 2$, and under Assumptions I and C1-C4, as $N, T \to \infty$:*

(a) $\hat{\boldsymbol{\lambda}}_{ji}\left(\hat{\theta}\right) \xrightarrow{p} \hat{\mathbf{H}}_{jj}\left(\theta^0\right)' \boldsymbol{\lambda}_{ji}^0$;

(b) $\hat{\mathbf{f}}_t \xrightarrow{p} \left[\mathbb{I}_{1t}\left(\theta^0\right) \hat{\mathbf{H}}_{11}\left(\theta^0\right)^{-1} + \mathbb{I}_{2t}\left(\theta^0\right) \hat{\mathbf{H}}_{22}\left(\theta^0\right)^{-1}\right] \mathbf{f}_t^0$;

(c) $\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\boldsymbol{\lambda}}_{ji}\left(\hat{\theta}\right) - \hat{\mathbf{H}}_{jj}\left(\theta^0\right)' \boldsymbol{\lambda}_{ji}^0\right\|^2 \xrightarrow{p} 0$;

(d) $\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{\mathbf{f}}_t - \left[\mathbb{I}_{1t}\left(\theta^0\right) \hat{\mathbf{H}}_{11}\left(\theta^0\right)^{-1} + \mathbb{I}_{2t}\left(\theta^0\right) \hat{\mathbf{H}}_{22}\left(\theta^0\right)^{-1}\right] \mathbf{f}_t^0\right\|^2 \xrightarrow{p} 0.$

3.5 Convergence Rates

The following theorem states the convergence rates of the concentrated least squares estimator for the threshold θ^0 and of the principal components estimator for the loadings.

Theorem 3.4 *Under Assumptions I, C1-C4 and CR,*

$$N^{\alpha^0} T\left(\hat{\theta} - \theta^0\right) = O_p(1)$$

and

$$C_{NT}^2\left[\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\boldsymbol{\lambda}}_{ji}\left(\hat{\theta}\right) - \hat{\mathbf{H}}_{jj}\left(\theta^0\right)' \boldsymbol{\lambda}_{ji}^0\right\|^2\right] = O_p(1), \quad j = 1, 2.$$

Theorem 3.4 states the superconsistency of $\hat{\theta}$ as an estimator for θ^0: it extends to an infinite dimensional system the result in Chan (1993) seminal contribution. The convergence rate $N^{\alpha^0}T$ of $\hat{\theta}$ depends on the time series dimension T *and* the number of cross-sectional units N^{α^0} subject to threshold effect: the rate $N^{\alpha^0}T$ monotonically increases in α^0; since $0.5 < \alpha^0 \leq 1$ by Assumption I, then $\sqrt{N}T < N^{\alpha^0}T \leq NT$; $N^{\alpha^0}T$ is unknown since α^0 is unknown. The higher α^0, the stronger identification of (1) from a linear factor model, and the faster the convergence rate of $\hat{\theta}$ to θ^0: this shows the connection between identification and estimation. When $\alpha^0 = 1$, all cross-sectional units are subject to threshold effect and the convergence rate is NT. Theorem 3.4 implies that the principal components estimator for the loadings has the same convergence rate derived in Bai and Ng (2002) in the case of linear factor models: the estimator for the threshold therefore does not affect the estimator for the loadings. Corollary 3.2 below follows from Theorem 3.4.

Corollary 3.2 *Under Assumptions I, C1-C4 and CR,*

$$C_{NT}^2\left[\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{\mathbf{f}}_t-\left[\mathbb{I}_{1t}\left(\theta^0\right)\hat{\mathbf{H}}_{11}\left(\theta^0\right)^{-1}+\mathbb{I}_{2t}\left(\theta^0\right)\hat{\mathbf{H}}_{22}\left(\theta^0\right)^{-1}\right]\mathbf{f}_t^0\right\|^2\right]=O_p\left(1\right).$$

Corollary 3.2 shows that the convergence rate C_{NT} also applies to the principal components estimator for the factors; it also shows that the rotation induced by $\hat{\mathbf{f}}_t$ around $\mathbf{f}_t^0$ depends upon the regime. Corollary 3.2 justifies the robust Bai and Ng (2002) information criteria proposed in Section 4.

4 Determining the Number of Factors

We now consider the case in which the true number of factors R^0 in (1) (i.e., the true dimension of $\mathbf{f}_t^0$) no longer is known and has to be determined. Breitung and Eickmeier (2011) show that neglecting structural breaks in the factor loadings inflates the estimated number of factors. Given the analogy between factor models with structural instability and (1), the latter suffers from the same problem. We rely on Corollary 3.2 and suggest a simple way to robustify Bai and Ng (2002) selection criteria to account for the threshold effect.

Given (1) and for fixed number of factors R, the loss function in (2) generalizes to

$$S\left(\mathbf{\Lambda}^R,\mathbf{F}^R,\theta\right)=\left(NT\right)^{-1}\sum_{t=1}^{T}\left[\mathbf{x}_t-\mathbb{I}_{1t}\left(\theta\right)\mathbf{\Lambda}_1^R\mathbf{f}_t^R-\mathbb{I}_{2t}\left(\theta\right)\mathbf{\Lambda}_2^R\mathbf{f}_t^R\right]'\left[\mathbf{x}_t-\mathbb{I}_{1t}\left(\theta\right)\mathbf{\Lambda}_1^R\mathbf{f}_t^R-\mathbb{I}_{2t}\left(\theta\right)\mathbf{\Lambda}_2^R\mathbf{f}_t^R\right], \tag{12}$$

where $\mathbf{\Lambda}^R=\left(\mathbf{\Lambda}_1^R,\mathbf{\Lambda}_2^R\right)$, $\mathbf{F}^R=\left(\mathbf{f}_1^R,\ldots,\mathbf{f}_T^R\right)$, and where the superscript R denotes the dependence on the number of factors. The loss function in (12) depends on θ. From Theorem 3.4, it easily follows that for any *a priori* chosen number of factors $R=\bar{R}$ such that $\bar{R}\geq R^0$, the estimator $\hat{\theta}^{\bar{R}}$ for θ^0 is such that $N^{\alpha^0}T\left(\hat{\theta}^{\bar{R}}-\theta^0\right)=O_p\left(1\right)$, with $\hat{\theta}^{R_0}=\hat{\theta}$ (see Lemma A.9 in Appendix A.3): in practice, $\bar{R}$ may be chosen as discussed below. Given the convergence rate in Corollary 3.2, this naturally suggests generalizing Bai and Ng (2002) criteria by first setting $\theta=\hat{\theta}^{\bar{R}}$ in (12) to then select $\hat{R}$ factors within each mutually exclusive regime, and therefore $\left(\hat{R}+\hat{R}\right)$ factors in total.

Let $\hat{\mathbf{\Lambda}}^R\left(\theta\right)$ and $\hat{\mathbf{F}}^R\left(\theta\right)$ be the estimators for $\mathbf{\Lambda}^R$ and $\mathbf{F}^R$, respectively, for any θ. Given the loss function in (12), and following Bai and Ng (2002), we want penalty functions $g\left(N,T\right)$ to obtain criteria of the form

$$PC(R, R) = S\left[\hat{\mathbf{\Lambda}}^{R}\left(\hat{\theta}^{\bar{R}}\right), \hat{\mathbf{F}}^{R}\left(\hat{\theta}^{\bar{R}}\right), \hat{\theta}^{\bar{R}}\right] + (R + R) \cdot g(N, T),$$

which consistently estimate the number of factors R^0 in each regime and therefore $\left(R^0 + R^0\right)$ factors in total: the criterion $PC(R, R)$ accounts for the fact that the threshold effect leads to a factor representation with a higher dimensional factor space, namely to a representation with $\left(R^0 + R^0\right)$ factors. Given a bounded integer $R^{\max} \geq R^0$, the true number of factors R^0 is estimated as

$$\hat{R} = \arg \min_{1 \leq R \leq R^{\max}} PC(R, R):$$

given the convergence rate C_{NT} in Corollary 3.2, this leads to the threshold effect robust Bai and Ng (2002) information criteria

$$IC_{p1}(R, R) = \ln S\left[\hat{\mathbf{\Lambda}}^{R}\left(\hat{\theta}^{\bar{R}}\right), \hat{\mathbf{F}}^{R}\left(\hat{\theta}^{\bar{R}}\right), \hat{\theta}^{\bar{R}}\right] + (R + R)\left(\frac{N+T}{NT}\right) \ln\left(\frac{NT}{N+T}\right),$$

$$IC_{p2}(R, R) = \ln S\left[\hat{\mathbf{\Lambda}}^{R}\left(\hat{\theta}^{\bar{R}}\right), \hat{\mathbf{F}}^{R}\left(\hat{\theta}^{\bar{R}}\right), \hat{\theta}^{\bar{R}}\right] + (R + R)\left(\frac{N+T}{NT}\right) \ln\left(C_{NT}^2\right), \tag{13}$$

$$IC_{p3}(R, R) = \ln S\left[\hat{\mathbf{\Lambda}}^{R}\left(\hat{\theta}^{\bar{R}}\right), \hat{\mathbf{F}}^{R}\left(\hat{\theta}^{\bar{R}}\right), \hat{\theta}^{\bar{R}}\right] + (R + R)\left[\frac{\ln\left(C_{NT}^2\right)}{C_{NT}^2}\right].$$

In practice, to obtain the estimator $\hat{\theta}^{\bar{R}}$ for θ^0, we may set $\bar{R} = R^{\max}$. The following theorem states the validity of the proposed information criteria.

Theorem 4.1 *Under Assumptions I, C1-C4 and CR, the criteria* $IC_{p1}(R, R)$*,* $IC_{p2}(R, R)$ *and* $IC_{p3}(R, R)$ *defined in* (13) *consistently estimate the number of factors* R^0*.*

The information criteria in (13) may be generalized by introducing a tuning multiplicative constant in the penalty as proposed in Alessi *et al.* (2010), who followed an idea put forward in Hallin and Liška (2007): it is high in our agenda to investigate the likely potential benefits of this method.

5 Testing for Linearity

5.1 Strategy and Test Statistic

Under Assumption I the model in (1) is identified from a linear factor model. We now extend Hansen (1996) seminal contribution to formally assess the validity of Assumption I.

Assumption LT1 - Linear Factor Model. $\sum_{i=N^{0.5}+1}^{N} \boldsymbol{\delta}_i^0 = O(1)$.

Under Assumption LT1, *no more* than $O\left(N^{0.5}\right)$ series undergo a regime shift. From Theorem 3.1, Assumption LT1 is the null hypothesis of linearity; Assumption I is the alternative. There exist several tests to detect structural breaks in large dimensional factor models: see Breitung and Eickmeier (2011), Han and Inoue (2015), and Yamamoto and Tanaka (2015). We follow Chen *et al.* (2014). Regime shifts in the loadings induce a change in the covariance matrix of the estimated factors. Let $\tilde{R}$ be the estimated number of factors in the linear model $\mathbf{x}_t = \boldsymbol{\Lambda}_1 \mathbf{f}_t + \mathbf{e}_t$: under Assumption LT1, $\tilde{R}$ is equal to the true number of factors, namely $\tilde{R} = R^0$; under Assumption I, $\tilde{R} = \left(R^0 + R^0\right)$ due to neglected regime shifts.

If $\tilde{R} = 1$ a regime shift in the loadings is ruled out with probability one. If $\tilde{R} > 1$ we proceed as follows. Let $\tilde{\mathbf{f}}_t$ be the $\tilde{R} \times 1$ vector of estimated factors from $\mathbf{x}_t = \boldsymbol{\Lambda}_1 \mathbf{f}_t + \mathbf{e}_t$, for $t = 1, \ldots, T$: consistently with Section 4, $\tilde{R}$ may be obtained as in Bai and Ng (2002). Following Chen *et al.* (2014), we construct the auxiliary threshold regression

$$\tilde{f}_{1t} = \mathbb{I}_{1t}(\theta)\, \boldsymbol{\beta}_1' \tilde{\mathbf{f}}_{-1,t} + \mathbb{I}_{2t}(\theta)\, \boldsymbol{\beta}_2' \tilde{\mathbf{f}}_{-1,t} + u_t, \quad t = 1, \ldots, T, \tag{14}$$

where $\tilde{f}_{1t} \in \mathfrak{R}$ is the first element of $\tilde{\mathbf{f}}_t$; $\tilde{\mathbf{f}}_{-1,t} \in \mathfrak{R}^{\tilde{R}-1}$ is the $\left(\tilde{R} - 1\right) \times 1$ vector containing the remaining elements of $\tilde{\mathbf{f}}_t$; $u_t \in \mathfrak{R}$ is the error term; $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are $\left(\tilde{R} - 1\right) \times 1$ vectors of slope coefficients. We test Assumption LT1 in (1) by testing $\boldsymbol{\beta}_1^0 = \boldsymbol{\beta}_2^0$ in (14), where $\boldsymbol{\beta}_1^0$ and $\boldsymbol{\beta}_2^0$ are the true values of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, respectively. This requires ruling out regime shifts in the covariance matrix of the factors. Let $\pi^0 = \mathrm{E}\left[\mathbb{I}_{1t}\left(\theta^0\right)\right]$ and recall $\boldsymbol{\Sigma}_{j\mathbf{f}}^0\left(\theta, \theta^0\right)$ in Assumption C1, for $j = 1, 2$.

Assumption LT2 - Threshold Effect in Factors. $T^{-1} \sum_{t=1}^{T} \mathbf{f}_t^0 \mathbf{f}_t^{0\prime} \xrightarrow{p} \boldsymbol{\Sigma}_{\mathbf{f}}^0$ as $T \to \infty$, $\boldsymbol{\Sigma}_{1\mathbf{f}}^0\left(\theta^0, \theta^0\right) = \pi^0 \boldsymbol{\Sigma}_{\mathbf{f}}^0$ and $\boldsymbol{\Sigma}_{2\mathbf{f}}^0\left(\theta^0, \theta^0\right) = \left(1 - \pi^0\right) \boldsymbol{\Sigma}_{\mathbf{f}}^0$, where $\boldsymbol{\Sigma}_{\mathbf{f}}^0$ is a positive definite matrix.

Assumption LT2 is analogous to Assumption 2 in Chen *et al.* (2014): if it fails to hold, the covariance matrix of the factors depends on the regimes and the test erroneously rejects the null hypothesis.

We build a Lagrange multiplier statistic (Hansen (1996)). Under Assumption LT1 the auxiliary regression in (14) reduces to $\tilde{f}_{1t} = \boldsymbol{\beta}_1' \tilde{\mathbf{f}}_{-1,t} + u_t$. The estimated factors are orthogonal to each other and $\tilde{f}_{1t} = u_t$: under the null hypothesis, the idiosyncratic component in (14) is generally serially correlated. Define $\tilde{\mathbf{f}}_{-,t}(\theta) = \left[\mathbb{I}_{1t}(\theta)\, \tilde{\mathbf{f}}_{-1,t}', \mathbb{I}_{2t}(\theta)\, \tilde{\mathbf{f}}_{-1,t}'\right]'$. For given θ, consider the estimator for $\boldsymbol{\beta}^0 = \left(\boldsymbol{\beta}_1^{0\prime}, \boldsymbol{\beta}_2^{0\prime}\right)'$

$$\hat{\boldsymbol{\beta}}(\theta) = \left[\hat{\boldsymbol{\beta}}_1(\theta)', \hat{\boldsymbol{\beta}}_2(\theta)'\right]' = \left[\sum_{t=1}^{T} \tilde{\mathbf{f}}_{-,t}(\theta)\, \tilde{\mathbf{f}}_{-,t}(\theta)'\right]^{-1} \left[\sum_{t=1}^{T} \tilde{\mathbf{f}}_{-,t}(\theta)\, \tilde{f}_{1t}\right].$$

For any (θ_1, θ_2), define the matrix $\hat{\mathbf{M}}_-(\theta_1, \theta_2) = T^{-1} \sum_{t=1}^{T} \tilde{\mathbf{f}}_{-,t}(\theta_1)\, \tilde{\mathbf{f}}_{-,t}(\theta_2)'$. The regression scores $\mathbf{k}_{-,t}(\theta) = \tilde{\mathbf{f}}_{-,t}(\theta)\, u_t$ are estimated under the null hypothesis as $\tilde{\mathbf{k}}_{-,t}(\theta) = \tilde{\mathbf{f}}_{-,t}(\theta)\, \tilde{f}_{1t}$. From Newey and West (1987), define: $\hat{\mathbf{K}}_{-,d}(\theta_1, \theta_2) = T^{-1} \sum_{t=d+1}^{T} \tilde{\mathbf{k}}_{-,t}(\theta_1)\, \tilde{\mathbf{k}}_{-,t-d}(\theta_2)'$, for $d = 0, \ldots, D_T$, with $D_T = o\left(T^{1/4}\right)$; $\hat{\Omega}_-(\theta_1, \theta_2) = \hat{\mathbf{K}}_{-,0}(\theta_1, \theta_2) + \sum_{d=1}^{D_T} w(d, D_T) \left[\hat{\mathbf{K}}_{-,d}(\theta_1, \theta_2) + \hat{\mathbf{K}}_{-,d}(\theta_1, \theta_2)'\right]$, where $w(d, D_T) = [1 - d/(D_T + 1)]$ is the Bartlett kernel. Define $\mathbf{G} = \left(\mathbf{I}_{\tilde{R}-1}, -\mathbf{I}_{\tilde{R}-1}\right)'$. For given θ, the heteroskedasticity and autocorrelation (HAC) robust Lagrange multiplier test statistic is

$$\widehat{LM}^{\mathrm{HAC}}(\theta) = T \hat{\boldsymbol{\beta}}(\theta)' \mathbf{G} \left[\mathbf{G}' \hat{\mathbf{M}}_-(\theta, \theta)^{-1} \hat{\Omega}_-(\theta, \theta)\, \hat{\mathbf{M}}_-(\theta, \theta)^{-1} \mathbf{G}\right]^{-1} \mathbf{G}' \hat{\boldsymbol{\beta}}(\theta).$$

For known θ^0 and under the null hypothesis, $\widehat{LM}^{\mathrm{HAC}}\left(\theta^0\right)$ has a χ^2 limiting distribution with $\left(R^0 - 1\right)$ degrees of freedom as $N, T \to \infty$. However, θ^0 is generally unknown and not identified under the null hypothesis. Following Davies (1977, 1987), and as in Hansen (1996), we propose the statistic

$$\sup \widehat{LM}^{\mathrm{HAC}} = \sup_{\theta} \widehat{LM}^{\mathrm{HAC}}(\theta). \tag{15}$$

When factors are serially uncorrelated, it is easy to show that (15) can be simplified to

$$\sup \widehat{LM}^{\mathrm{HC}} = \sup_{\theta} \widehat{LM}^{\mathrm{HC}}(\theta), \tag{16}$$

with

$$\widehat{LM}^{\mathrm{HC}}(\theta) = T \hat{\boldsymbol{\beta}}(\theta)' \mathbf{G} \left[\mathbf{G}' \hat{\mathbf{M}}_-(\theta, \theta)^{-1} \hat{\mathbf{K}}_{-,0}(\theta, \theta)\, \hat{\mathbf{M}}_-(\theta, \theta)^{-1} \mathbf{G}\right]^{-1} \mathbf{G}' \hat{\boldsymbol{\beta}}(\theta).$$

The heteroskedasticity robust statistic $\sup \widehat{LM}^{\mathrm{HC}}$ in (16) is analogous to the one studied in Hansen (1996): we construct the more general heteroskedasticity and autocorrelation robust statistic in (15).

5.2 Limiting Distribution under the Null Hypothesis

Let $\hat{\mathbf{k}}_{-}(\theta) = T^{-1/2}\sum_{t=1}^{T}\tilde{\mathbf{k}}_{-,t}(\theta)$ and $\mathbf{k}_{-}^{0}(\theta)$ be a zero mean Gaussian process with covariance kernel $\Omega_{-}^{0}(\theta_1,\theta_2) = \mathrm{E}\left[\mathbf{k}_{-}^{0}(\theta_1)\mathbf{k}_{-}^{0}(\theta_2)'\right]$. Define $\hat{\mathbf{M}}(\theta_1,\theta_2) = T^{-1}\sum_{t=1}^{T}\left[\mathbb{I}_{1t}(\theta_1)\mathbf{f}_t^{0\prime},\mathbb{I}_{2t}(\theta_1)\mathbf{f}_t^{0\prime}\right]'\left[\mathbb{I}_{1t}(\theta_2)\mathbf{f}_t^{0\prime},\mathbb{I}_{2t}(\theta_2)\mathbf{f}_t^{0\prime}\right]$ and $\mathbf{M}^{0}(\theta_1,\theta_2) = \mathrm{E}\left\{\left[\mathbb{I}_{1t}(\theta_1)\mathbf{f}_t^{0\prime},\mathbb{I}_{2t}(\theta_1)\mathbf{f}_t^{0\prime}\right]'\left[\mathbb{I}_{1t}(\theta_2)\mathbf{f}_t^{0\prime},\mathbb{I}_{2t}(\theta_2)\mathbf{f}_t^{0\prime}\right]\right\}$ under Assumption LT5(a) below.

Assumption LT3 - Eigenvalues. The eigenvalues of the $R^0 \times R^0$ matrix $\left(\mathbf{\Sigma}_{\mathbf{f}}^{0}\cdot\mathbf{D}_{\mathbf{\Lambda}_1}^{0}\right)$ are distinct.

Assumption LT4 - Convergence Rates. $\sqrt{T}/N \to 0$ as $N\to\infty$ and $T\to\infty$.

Assumption LT5 - Mixing Condition and Moment Bound.

(a) $\left\{\mathbf{f}_t^0, z_t\right\}_{t=1}^{T}$ is strictly stationary and $\beta-$mixing, with $\beta-$mixing coefficients satisfying $\beta_m = O\left(m^{-\nu}\right)$ for some $\nu > \xi/(\xi-1)$ and $r \geq \xi > 1$;

(b) $\mathrm{E}\left\{\left|\max_{j=1,2}\left[\sup_{\theta}\left\|\mathbb{I}_{jt}(\theta)\mathbf{f}_t^{0}\right\|\right]\right|^{4r}\right\} < \infty$.

Assumption LT6 - Bracketing. For all θ, and for some $M<\infty$ and $\gamma>0$, there exists some $\bar{\theta}$ such that $\left\{\mathrm{E}\left|\max_{j=1,2}\left\|\left[\mathbb{I}_{jt}(\theta)-\mathbb{I}_{jt}\left(\bar{\theta}\right)\right]\mathbf{f}_t^{0}\mathbf{f}_t^{0\prime}\right\|\right|^{2\xi}\right\}^{1/(2\xi)} \leq M\left|\theta-\bar{\theta}\right|^{\gamma}$.

Assumption LT7 - Uniform Convergence. $\hat{\mathbf{M}}(\theta_1,\theta_2)$ and $\hat{\Omega}_{-}(\theta_1,\theta_2)$ converge in probability to $\mathbf{M}^{0}(\theta_1,\theta_2)$ and $\Omega_{-}^{0}(\theta_1,\theta_2)$, respectively, uniformly over (θ_1,θ_2), where $\mathbf{M}^{0}(\theta_1,\theta_2)$ and $\Omega_{-}^{0}(\theta_1,\theta_2)$ are positive definite matrices.

Assumption LT3 is analogous to Assumption G in Bai (2003) and guarantees a unique probability limit for $\left(\mathbf{\Lambda}_1^{0\prime}\tilde{\mathbf{\Lambda}}_1/N\right)$. Assumption LT4 imposes a standard restriction on the convergence rates. Assumptions LT5-LT7 are equivalent to Assumptions 1-3 in Hansen (1996), respectively. The uniform convergence of $\hat{\Omega}_{-}(\theta_1,\theta_2)$ to $\Omega_{-}^{0}(\theta_1,\theta_2)$ is not stringent: factors are consistently estimated under Assumptions C1-C4, LT1 and LT3; and $\hat{\Omega}_{-}(\theta_1,\theta_2)$ is a HAC estimator for the covariance kernel $\Omega_{-}^{0}(\theta_1,\theta_2)$ (see also Assumption 11 in Chen *et al.* (2014)). Assumptions LT5 and LT7 jointly imply Assumption C1.

Let $\mathbf{M}_{-}^{0}(\theta_1,\theta_2)$ be such that $\hat{\mathbf{M}}_{-}(\theta_1,\theta_2)\xrightarrow{p}\mathbf{M}_{-}^{0}(\theta_1,\theta_2)$ for any (θ_1,θ_2) as $N,T\to\infty$: the existence of $\mathbf{M}_{-}^{0}(\theta_1,\theta_2)$ is guaranteed by Assumption LT7. Define

$$LM^{\mathrm{HAC},0}(\theta) = \left[\mathbf{M}_{-}^{0}(\theta,\theta)^{-1}\mathbf{k}_{-}^{0}(\theta)\right]'\mathbf{G}\left[\mathbf{G}'\mathbf{M}_{-}^{0}(\theta,\theta)^{-1}\Omega_{-}^{0}(\theta,\theta)\mathbf{M}_{-}^{0}(\theta,\theta)^{-1}\mathbf{G}\right]^{-1}\mathbf{G}'\left[\mathbf{M}_{-}^{0}(\theta,\theta)^{-1}\mathbf{k}_{-}^{0}(\theta)\right]$$

and

$$\sup LM^{\text{HAC},0} = \sup_{\theta} LM^{\text{HAC},0}(\theta).$$

Theorem 5.1 *Under Assumptions C2-C4 and LT1-LT7,* $\hat{\mathbf{k}}_{-}(\theta) \Rightarrow \mathbf{k}_{-}^{0}(\theta)$, $\widehat{LM}^{\text{HAC}}(\theta) \Rightarrow LM^{\text{HAC},0}(\theta)$, *and* $\sup \widehat{LM}^{\text{HAC}} \xrightarrow{d} \sup LM^{\text{HAC},0}$, *as* $N, T \to \infty$.

Theorem 5.1 implies that Hansen (1996) fixed regressor bootstrap approximates the asymptotic distribution of $\sup \widehat{LM}^{\text{HAC}}$ in (15) under the null hypothesis[4]. For $b = 1, \ldots, \bar{b}$: (i) generate $u_{bt}^{*} \sim \text{IID}\mathcal{N}(0,1)$; (ii) define $\hat{\mathbf{k}}_{-,b}^{*}(\theta) = T^{-1/2}\sum_{t=1}^{T} \tilde{\mathbf{f}}_{-,t}(\theta)\tilde{f}_{1t}u_{bt}^{*}$; (iii) let $\sup \widehat{LM}_{b}^{\text{HAC},*} = \sup_{\theta} \widehat{LM}_{b}^{\text{HAC},*}(\theta)$, where

$$\widehat{LM}_{b}^{\text{HAC},*}(\theta) = \left[\hat{\mathbf{M}}_{-}(\theta,\theta)^{-1}\hat{\mathbf{k}}_{-,b}^{*}(\theta)\right]' \mathbf{G}\left[\mathbf{G}'\hat{\mathbf{M}}_{-}(\theta,\theta)^{-1}\hat{\Omega}_{-}(\theta,\theta)\hat{\mathbf{M}}_{-}(\theta,\theta)^{-1}\mathbf{G}\right]^{-1}\mathbf{G}'\left[\hat{\mathbf{M}}_{-}(\theta,\theta)^{-1}\hat{\mathbf{k}}_{-,b}^{*}(\theta)\right].$$

The empirical distribution of $\left\{\sup \widehat{LM}_{b}^{\text{HAC},*}\right\}_{b=1}^{\bar{b}}$ approximates the asymptotic distribution of $\sup \widehat{LM}^{\text{HAC}}$ under the null hypothesis of linearity as stated in Assumption LT1.

6 Monte Carlo Analysis

The experiments related to estimation, model selection and linearity testing are described in Sections 6.1, 6.2 and 6.3, respectively; the results are discussed in Section 6.4.

6.1 Estimation

In line with the results in Section 3, we assume a known number of factors. As in Breitung and Eickmeier (2011), we analyze a one-factor model. We simulate the data using the Data Generating Process (DGP)

$$x_{it}^{s} = \mathbb{I}\left(z_{t}^{s} \leq \theta^{0}\right)\lambda_{1i}^{0}f_{t}^{0s} + \mathbb{I}\left(z_{t}^{s} > \theta^{0}\right)\lambda_{2i}^{0}f_{t}^{0s} + e_{it}^{s}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T,$$

where $s = 1, \ldots, S$ refers to the replication and S is the total number of replications. We set $S = 2000$, $N = 25, 50, 100$ and $T = 100, 200, 400$. We define $\delta_{i}^{0} = \lambda_{2i}^{0} - \lambda_{1i}^{0}$: we set $\delta_{i}^{0} > 0$ for $i = 1, \ldots, \left[N^{\alpha^{0}}\right]$ and $\delta_{i}^{0} = 0$ for $i = \left[N^{\alpha^{0}}\right] + 1, \ldots, N$, where $[\cdot]$ denotes the integer part of the argument. We fix the factor loadings λ_{1i}^{0} and λ_{2i}^{0} and the threshold parameter θ^{0} throughout the replications, with $\lambda_{1i}^{0} \sim \mathcal{N}(1,1)$ for $i = 1, \ldots, N$ as in the Monte Carlo experiment in Breitung and Eickmeier (2011), and $\theta^{0} = 2$. We control

[4] A formal proof would follow similar steps as that of Theorem 2 in Hansen (1996) and it is omitted.

for: (i) the number of cross-sectional units $\left[N^{\alpha^0}\right]$ subject to a regime change by setting $\alpha^0 = 0.60, 1.00$; and (ii) the magnitude of the threshold effect by setting $\delta_i^0 = 0.25, 1.00, 1.75$. We generate z_t^s as

$$z_t^s = \mu_z (1 - \rho_z) + \rho_z z_{t-1}^s + \left(1 - \rho_z^2\right)^{1/2} \epsilon_{zt}^s, \quad z_{-50}^s = \mu_z, \quad t = -49, \dots, 0, \dots, T, \tag{17}$$

where μ_z and $\rho_z \sim \mathcal{U}(0.05, 0.95)$ are fixed in repeated samples, and $\epsilon_{zt}^s \sim \mathrm{IID}\mathcal{N}(0,1)$: in this way $\mathrm{E}(z_t^s) = \mu_z$ and $\mathrm{Var}(z_t^s) = 1$. We let $\pi^0 = \mathrm{P}\left(z_t^s \leq \theta^0\right) = \mathrm{P}\left(z_t^s - \mu_z \leq \theta^0 - \mu_z\right) = \mathbf{\Phi}\left(\theta^0 - \mu_z\right) = 0.50$ and obtain $\mu_z = \theta^0 - \mathbf{\Phi}^{-1}\left(\pi^0\right) = 2$: the choice $\pi^0 = 0.50$ is consistent with the existing literature (see Breitung and Eickmeier (2011), Chen *et al.* (2014), and Han and Inoue (2015)).

We generate the factor f_t^{0s} as

$$f_t^{0s} = \rho_f f_{t-1}^{0s} + \left(1 - \rho_f^2\right)^{1/2} \varpi_{ft}^s \epsilon_{ft}^s, \quad f_{-50}^{0s} = 0, \quad t = -49, \dots, 0, \dots, T, \tag{18}$$

with $\rho_f \sim \mathcal{U}(0.05, 0.95)$ fixed in repeated samples, $\mathrm{E}\left(\varpi_{ft}^s\right)^2 = 1$ and $\epsilon_{ft}^s \sim \mathrm{IID}\mathcal{N}(0,1)$, so $\mathrm{E}\left(f_t^{0s}\right) = 0$ and $\mathrm{Var}\left(f_t^{0s}\right) = 1$. We allow for conditional heteroskedasticity in f_t^{0s} through the GARCH$(1,1)$ process $\left(\varpi_{ft}^s\right)^2 = \beta_{f1} + \beta_{f2}\left(\varpi_{f,t-1}^s\right)^2 + \beta_{f3}\left(\varpi_{f,t-1}^s \epsilon_{f,t-1}^s\right)^2$, with $\left(\varpi_{f,-50}^s\right)^2 = \mathrm{E}\left(\varpi_{ft}^s\right)^2 = 1$.

We generate the idiosyncratic components e_{it}^s as

$$e_{it}^s = \rho_e e_{i,t-1}^s + \sigma_{ii}^{1/2}\left(1 - \rho_e^2\right)^{1/2} \varpi_{e_i t}^s \epsilon_{e_i t}^s, \quad e_{i,-50}^s = 0, \quad i = 1, \dots, N, \quad t = -49, \dots, 0, \dots, T, \tag{19}$$

with $\rho_e \sim \mathcal{U}(0.05, 0.95)$ and $\sigma_{ii} \sim \chi(1)$ fixed in repeated samples. Let $\boldsymbol{\epsilon}_{et}^s = \left(\epsilon_{e_1 t}^s, \dots, \epsilon_{e_N t}^s\right)'$. We allow for cross-sectional dependence through the first order spatial autoregressive process $\boldsymbol{\epsilon}_{et}^s = \bar{\mathbf{Q}} \boldsymbol{\varrho}_{et}^s$, where $\bar{\mathbf{Q}} = \mathbf{Q}\left[N \Big/ \mathrm{tr}\left(\mathbf{\Sigma}_{e,\mathrm{diag}}^{1/2} \mathbf{\Omega}_{e,\mathrm{diag}}^{1/2} \mathbf{Q}\mathbf{Q}' \mathbf{\Omega}_{e,\mathrm{diag}}^{1/2} \mathbf{\Sigma}_{e,\mathrm{diag}}^{1/2}\right)\right]^{1/2}$, $\mathbf{\Sigma}_{e,\mathrm{diag}}^{1/2} = \mathrm{diag}\left[\left(\sigma_{11}^{1/2}, \dots, \sigma_{NN}^{1/2}\right)'\right]$, $\mathbf{\Omega}_{e,\mathrm{diag}}^{1/2} = \mathrm{diag}\left\{\left\{\left[\mathrm{E}\left(\varpi_{e_1 t}^s\right)^2\right]^{1/2}, \dots, \left[\mathrm{E}\left(\varpi_{e_N t}^s\right)^2\right]^{1/2}\right\}'\right\}$, $\boldsymbol{\varrho}_{et}^s \sim \mathrm{IID}\mathcal{N}(0, \mathbf{I}_N)$, and $\mathbf{Q} = \left(\mathbf{I}_N - \iota \mathbf{W}\right)^{-1}$ with

$$\mathbf{W} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0.5 & 0 & 0.5 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0.5 \\ 0 & \dots & \dots & 1 & 0 \end{pmatrix}:$$

in this way $\operatorname{Var}(e_{it}^s) = \sigma_{ii}\mathrm{E}\left(\varpi_{e_it}^s\right)^2 \Big/ \left[N^{-1}\sum_{l=1}^N \sigma_{ll}\mathrm{E}\left(\varpi_{e_lt}^s\right)^2\right]$ and $N^{-1}\sum_{i=1}^N \operatorname{Var}(e_{it}^s) = 1$. We model $\varpi_{e_it}^s$ as the GARCH$(1,1)$ process $\left(\varpi_{e_it}^s\right)^2 = \beta_{e1} + \beta_{e2}\left(\varpi_{e_i,t-1}^s\right)^2 + \beta_{e3}\left(\varpi_{e_i,t-1}^s \epsilon_{e_i,t-1}^s\right)^2$, with $\left(\varpi_{e_i,-50}^s\right)^2 = \mathrm{E}\left(\varpi_{e_it}^s\right)^2 = 1$: it follows that $\operatorname{Var}(e_{it}^s) \to \sigma_{ii}$ as $N \to \infty$.

We consider three scenarios: (i) time homoskedastic factors and idiosyncratic components, and cross-sectionally independent idiosyncratic components (CSI); (ii) time homoskedastic factors and idiosyncratic components, and cross-sectionally dependent idiosyncratic components (CSD); and (iii) time heteroskedastic factors and idiosyncratic components, and cross-sectionally dependent idiosyncratic components (CSDH). Under CSI, we set $\beta_{f1} = \beta_{e1} = 1$, $\beta_{f2} = \beta_{e2} = 0$, $\beta_{f3} = \beta_{e3} = 0$ and $\iota = 0$. We build CSD by imposing $\beta_{f1} = \beta_{e1} = 1$, $\beta_{f2} = \beta_{e2} = 0$, $\beta_{f3} = \beta_{e3} = 0$ and $\iota = 0.4$. We parameterize CSDH by setting $\beta_{f1} = \beta_{e1} = 0.1$, $\beta_{f2} = \beta_{e2} = 0.8$, $\beta_{f3} = \beta_{e3} = 0.1$ and $\iota = 0.4$.

To reduce the effect induced by the initial values $z_{-50}^s = \mu_z$, $f_{-50}^{0s} = 0$, $\varpi_{f,-50}^s = 1$, $e_{i,-50}^s = 0$ and $\varpi_{e_i,-50}^s = 1$, we discard the first 50 observations in the DGPs for z_t^s, f_t^{0s}, ϖ_{ft}^s, e_{it}^s and $\varpi_{e_it}^s$. We estimate factor and loadings as detailed in Section 3.3. Given the convergence rates Theorem 3.4, the estimator for θ^0 is asymptotically independent of that for λ_{1i}^0, λ_{2i}^0 and f_t^{0s}. As in Tong and Lim (1980), Tsay (1989) and Kapetanios (2000), we estimate θ^0 by grid search: we implement the algorithm by selecting 19 equally spaced quantiles of the empirical distribution function of z_t^s, namely $\{5\%, 10\%, 15\%, \ldots, 85\%, 90\%, 95\%\}$, and the true value $\theta^0 = 2$. Given the concentrated least squares estimator $\hat{\theta}^s$ for θ^0, we estimate factor and loadings by principal components. We assess $\hat{\theta}^s$ by computing

$$\text{bias} = S^{-1}\sum_{s=1}^S \left(\hat{\theta}^s - \theta^0\right), \quad \text{RMSE} = \sqrt{S^{-1}\sum_{s=1}^S \left(\hat{\theta}^s - \theta^0\right)^2}.$$

Finally, given the estimator $\hat{c}_{it}^s = \mathbb{I}\left(z_t^s \le \hat{\theta}^s\right)\hat{\lambda}_{1i}^s \hat{f}_t^s + \mathbb{I}\left(z_t^s > \hat{\theta}^s\right)\hat{\lambda}_{2i}^s \hat{f}_t^s$ for the common component $c_{it}^{0s} = \mathbb{I}\left(z_t^s \le \theta^0\right)\lambda_{1i}^0 f_t^{0s} + \mathbb{I}\left(z_t^s > \theta^0\right)\lambda_{2i}^0 f_t^{0s}$, we report

$$\text{MSE} = S^{-1}\sum_{s=1}^S \left[(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T \left(\hat{c}_{it}^s - c_{it}^{0s}\right)^2\right].$$

6.2 Model Selection

We simulate the data using the two-factor DGP

$$x_{it}^{s}=\mathbb{I}\left(z_{t}^{s}\leq\theta^{0}\right)\left(\lambda_{11i}^{0}f_{1t}^{0s}+\lambda_{12i}^{0}f_{2t}^{0s}\right)+\mathbb{I}\left(z_{t}^{s}>\theta^{0}\right)\left(\lambda_{21i}^{0}f_{1t}^{0s}+\lambda_{22i}^{0}f_{2t}^{0s}\right)+e_{it}^{s},\quad i=1,\ldots,N,\quad t=1,\ldots,T,$$

with $\lambda_{11i}^{0}\sim\mathcal{N}(1,1)$, $\lambda_{12i}^{0}\sim\mathcal{N}(1,1)$, $\lambda_{21i}^{0}=\lambda_{11i}^{0}+\delta_{i}^{0}$ and $\lambda_{22i}^{0}=\lambda_{12i}^{0}+\delta_{i}^{0}$. We set $\delta_{i}^{0}=0.25,1.00,1.75$ for $i=1,\ldots,[N^{\alpha_0}]$, and $\delta_{i}^{0}=0$ for $i=[N^{\alpha_0}]+1,\ldots,N$, with $\alpha_0=0.60$. The factors f_{1t}^{0s} and f_{2t}^{0s} are generated as AR(1) processes analogous to (18); z_t^s and e_{it}^s are as in (17) and (19), respectively. The model has $R^0=2$ factors and it is estimated with $R^{\max}=8$. We assess the model selection criteria in (13) by reporting the average number of estimated factors over the 2000 replications.

6.3 Linearity Testing

Under the null hypothesis, we simulate the data from the linear two-factor model

$$x_{it}^{s}=\lambda_{1i}^{0}f_{1t}^{0s}+\lambda_{2i}^{0}f_{2t}^{0s}+e_{it}^{s},\quad i=1,\ldots,N,\quad t=1,\ldots,T,$$

with $\lambda_{1i}^{0}\sim\mathcal{N}(1,1)$, $\lambda_{2i}^{0}\sim\mathcal{N}(1,1)$. The factors f_{1t}^{0s} and f_{2t}^{0s} are generated as AR(1) processes analogous to (18) and we look at two cases: (i) $\rho_f=0$, factors are serially uncorrelated and the heteroskedasticity robust statistic in (16) is used; and (ii) $\rho_f=0.5$, factors have time dependence and the HAC statistic in (15) is used with Barlett window $D_T=5$. Under the alternative hypothesis, we simulate the data from the one-factor model in Section 6.1, with $\alpha^0=0.60$: we set $\rho_f=0.5$ in (18), factors are serially correlated and the HAC statistic in (15) is used. We set the number of bootstrap replications to $\bar{b}=1000$.

6.4 Results

The results are collected in four tables: Tables 1 and 2 focus on estimation; model selection criteria are assessed in Table 3; size and power of the linearity test are shown in Table 4.

Table 1 about here

Table 2 about here

Table 3 about here

Table 4 about here

Table 1 displays results for the concentrated least squares estimator $\hat{\theta}$ for $\theta^0 = 2$ when $\alpha^0 = 0.60$ (Panel A) and $\alpha^0 = 1.00$ (Panel B). Given Theorems 3.1 and 3.4, a higher α^0 leads to stronger identification of θ^0 and faster convergence rate of $\hat{\theta}$ to θ^0, respectively: in line with these theoretical results, the RMSE of $\hat{\theta}$ when $\alpha^0 = 1.00$ is generally lower than the homologous value when $\alpha^0 = 0.60$ under CSI, CSD and CSDH. The RMSE tends to decrease with N, T and $\delta_i^0 > 0$. The RMSE also increases as cross-sectional dependence and time heteroskedasticity are added to the DGP as compared to the CSI scenario. The bias displays a pattern somehow similar to that of the RMSE.

Table 2 shows the MSE of the common components when $\alpha^0 = 0.60$ (Panels A) and $\alpha^0 = 1.00$ (Panels B). We assess the empirical validity of Theorem 3.4 by considering both unfeasible and feasible estimators, the former and the latter being obtained by setting $\theta = \theta^0$ and $\theta = \hat{\theta}$, respectively. In line with Theorem 3.4, the MSE of the feasible estimator converges to that of the unfeasible counterpart as both N and T increase. The MSE monotonically decreases in N and T, and in $\delta_i^0 > 0$ for $N = 25$, whereas it does not exhibit any systematically noticeable difference between $\alpha^0 = 0.60$ and $\alpha^0 = 1.00$. The MSE also increases when cross-sectional dependence is added to the DGP, whereas it seems to be less affected by time heteroskedasticity.

Table 3 collects results for the selection criteria $IC_{p1}(R, R)$, $IC_{p2}(R, R)$ and $IC_{p3}(R, R)$ (Panels A, B and C, respectively) in (13) when $\alpha^0 = 0.60$. The criteria $IC_{p1}(R, R)$ and $IC_{p2}(R, R)$ display a similar behavior under CSI, with the latter having a hedge over the former: they tend to overestimate the number of factors for $N = 25, 50$, whereas they perform well for $N = 100$. The criterion $IC_{p2}(R, R)$ is the best under both CSD and CSDH, where the performance of $IC_{p2}(R, R)$ slightly deteriorates as compared to CSI. The criterion $IC_{p3}(R, R)$ is the least accurate under all scenarios. Finally, unfeasible and feasible estimators give similar results in terms of model selection performance.

Finally, Table 4 reports results for the linearity test at 5% and 10% level (Panels A and B, respectively). Regardless of ρ_f and N, the test is correctly sized for $T = 400$. It is undersized for lower values of T, with the exception of scenario CSDH with $\rho_f = 0.00$ and $T = 200$. The test has size properties analogous to Breitung and Eickmeier (2011) Lagrange multiplier test under unknown break-point. The power increases in N, T and $\delta_i^0 > 0$, though the effect of size distortions ought to be taken into account.

In conclusion, the Monte Carlo findings corroborate the theoretical results stated in Theorems 3.1 and 3.4. They confirm the validity of the information criteria in (13) and suggest using $IC_{p2}(R, R)$. Finally, they show that the proposed linearity test is able to detect regime shifts.

7 Empirical Application

We show how our framework may be used to measure connectedness in multivariate nonlinear dynamic systems, with a focus on financial variables: a threshold factor specification is suitable when "history repeats", as in financial markets, which undergo regime shifts (Timmermann (2008), and Ang and Timmermann (2012)). Section 7.1 proposes a measure of connectedness, Section 7.2 describes the data and the empirical model, and Section 7.3 presents the results.

7.1 Measure of Connectedness

Connectedness is central to risk measurement and management. There exist several measures of connectedness, which are based on different underlying metrics: examples are the marginal expected shortfall of Acharya *et al.* (2010), the equicorrelation approach of Engle and Kelly (2012), the network approach of Diebold and Yilmaz (2014), and the CoVaR of Adrian and Brunnermeier (2016). In line with our methodological contribution, we focus on the principal components approach of Billio *et al.* (2012).

Given the sequence of $N \times 1$ vectors $\{\mathbf{x}_t\}_{t=1}^T$, let $\{\omega_r\}_{r=1}^N$ be the sequence of eigenvalues of the $N \times N$ covariance matrix $\hat{\mathbf{\Sigma}}_{\mathbf{x}} = (NT)^{-1} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'$. In relation to financial markets, Billio *et al.* (2012) quantify the degree of connectedness amongst the elements of $\mathbf{x}_t$ as the risk associated to the first R eigenvalues in relation to the overall risk of the system. Formally, they measure connectedness through[5]

$$C(R) = \frac{\sum_{r=1}^R \omega_r}{\sum_{r=1}^N \omega_r} :$$

by construction $C(R)$ is increasing in R; for given R, a higher $C(R)$ denotes higher connectedness amongst the underlying variables. The measure $C(R)$ powerfully captures connectedness amongst random variables. However, it suffers from two main drawbacks. First, the number of eigenvalues R is

[5] Billio *et al.* (2012) refer to $C(R)$ as to the Cumulative Risk Fraction.

chosen *a priori* and not according to a selection criterion. Second, $C(R)$ refers to the entire time series dimension T and is unable to detect variations in connectedness induced by a threshold effect. Financial markets experience regimes shifts (Timmermann (2008), and Ang and Timmermann (2012)): the measure $C(R)$ may not accurately describe the dynamics in connectedness of the variables of interest[6]. Our methodology allows to build a connectedness measure that accommodates regime shifts and relies on the optimally selected number of eigenvalues.

Let $\{\omega_{jr}\}_{r=1}^{N}$ be the sequence of eigenvalues of the $N \times N$ covariance matrix $\hat{\mathbf{\Sigma}}_{j\mathbf{x}}(\theta)$ defined in (7) in decreasing order, for $j = 1, 2$. We generalize $C(R)$ and measure connectedness through

$$C_j\left(\hat{R}\right) = \frac{\sum_{r=1}^{\hat{R}} \omega_{jr}}{\sum_{r=1}^{N} \omega_{jr}}, \quad j = 1, 2. \tag{20}$$

Compared to $C(R)$, the measure $C_j\left(\hat{R}\right)$ has two distinctive features: it quantifies connectedness within each regime; and the number of eigenvalues $\hat{R}$ is optimally determined according to the criteria in (13).

7.2 Data and Model Specification

We construct the vector of dependent variables from the updated monthly financial dataset employed in Jurado *et al.* (2015) and, on a quarterly frequency, in Ludvigson and Ng (2007)[7]: this consists of a panel of 147 series related to the U.S. financial markets, as detailed in Ludvigson and Ng (2007).

We study how economic policy uncertainty affects connectedness amongst financial variables. The threshold variable is the lagged index of economic policy uncertainty proposed in Baker *et al.* (2016)[8]: a higher index value denotes higher uncertainty. Financial markets uncertainty leads to economic policy uncertainty and the threshold variable is likely to be predetermined (see discussion in Section 3.1.2).

Due to data availability issues, we perform the empirical analysis over the period running from January 1985 to December 2014, a total of 360 observations. The threshold variable has mean, standard deviation, maximum and minimum equal to 107.640, 32.566, 245.127 and 57.203, respectively.

We fit a linear factor model to the data and select 8 factors using the $IC_{p2}(R)$ criterion of Bai and Ng (2002). Neither $\sup \widehat{LM}^{\text{HAC}}$ nor $\sup \widehat{LM}^{\text{HC}}$ in (15) and (16), respectively, reject the null of linearity:

[6] Billio *et al.* (2012) measure the dynamic degree of connectedness in financial returns by computing $C(R)$ over rolling windows.

[7] I am very grateful to Sydney Ludvigson for providing me with the updated version of the dataset I am using in the paper. See Jurado *et al.* (2015) for a more detailed description of the data.

[8] The index is made available at http://www.policyuncertainty.com/ .

the tests are likely to have low power when applied to financial data, as market efficiency limits factors explanatory ability. As customary in empirical asset pricing, we still select two regimes (see Ang and Timmermann (2012)). We consider $R^{\max} = 10$ and estimate the change-point by setting $\bar{R} = R^{\max}$; we then construct a grid for the change-point with lowest and highest values equal to 5% and 95%, respectively, and step equal to 0.5%. The number of factors are selected according to the criteria in (13).

7.3 Results

Results are collected in Table 5.

Table 5 about here

The point estimate for the threshold θ^0 is $\hat{\theta} = 131.413$: this splits the sample into low and high economic policy uncertainty regimes, with frequencies equal to $\hat{\pi} = 0.783$ and $1 - \hat{\pi} = 0.217$, respectively. Figure 1 shows the high uncertainty regime, as identified by $\mathbb{I}\left(z_t > \hat{\theta}\right) = 1$, plotted against time.

Figure 1 about here

The criteria $IC_{p1}(R, R)$ and $IC_{p2}(R, R)$ select $\hat{R} = 3$ factors, with connectedness measures $C_1\left(\hat{R}\right) = 0.678$ and $C_2\left(\hat{R}\right) = 0.865$. Conversely, $IC_{p3}(R, R)$ selects $\hat{R}_1 = 6$ factors: this is consistent with the Monte Carlo results in Section 6.4, which show that $IC_{p3}(R, R)$ overestimates the number of factors in finite samples. Our results show that connectedness amongst financial variables increases with economic policy uncertainty: this likely to be relevant for risk measurement and management.

8 Directions for Future Research

We outline two directions for future research. It would be useful to apply to (1) the projected principal components estimator of Fan *et al.* (2016a). By including additional covariates in the information set, this would allow to consistently estimate factors and loadings without requiring $T \to \infty$: this would be important as the regimes in (1) effectively reduce the available time dimension.

Following Fan *et al.* (2013, 2016b), and Bai and Liao (2016), it would be interesting to introduce conditional sparsity in (1). Conditional sparsity allows to estimate the error covariance matrix in large dimensional approximate factor models by imposing that many entries are zero or nearly zero. In a linear framework, Fan *et al.* (2013) develop a two-step procedure that first estimates factors and loadings by principal components, and then applies a thresholding procedure to the remaining covariance matrix. Bai and Liao (2016) propose a penalized maximum likelihood method that jointly estimates loadings and error covariance matrix: the factors are then estimated by generalized least squares. Fan *et al.* (2016b) robustify Fan *et al.* (2013) estimator to account for asymmetric and heavy tailed error distribution. As applied to (1), conditional sparsity would have to be imposed within each regime: this would allow to estimate regime-specific error covariance matrices; from the superconsistency property in Theorem 3.4, the results in Fan *et al.* (2013, 2016b), and Bai and Liao (2016) would then apply within each regime.

9 Conclusions

We study least squares estimation of large dimensional factor models with threshold-type regime shifts in the loadings. Our methodology handles the general case of unknown threshold parameter. The concentrated least squares estimator for the threshold value is superconsistent: the convergence rate depends on the time series dimension and on the number of cross-sectional units subject to threshold effect. The principal components estimator for factors and loadings has the same convergence rate as in linear factor models: this allows to robustify Bai and Ng (2002) selection criteria by accounting for the higher dimensional factor space representation induced by the regime shift. We also propose a simple yet powerful linearity test to detect regime changes. In an application, we document an increase in connectedness amongst financial variables during periods of high economic policy uncertainty: this result is likely to be relevant for risk measurement and management.

A Proofs of Theorems

A.1 Proofs of Results in Section 3.4

We rely on the following lemmas.

Lemma A.1 *Under Assumptions I and C1-C3, there exists some positive constant $M < \infty$ such that for all θ, all (N,T) and $j = 1, 2$:*

(a) $N^{-1} \sum_{i=1}^{N} \sum_{l=1}^{N} \sigma_{jil}^{2}(\theta) \leq M$;

(b) $\mathrm{E}\left\{N^{-2} \sum_{i=1}^{N} \sum_{l=1}^{N} \left[T^{-1} \sum_{t=1}^{T} \mathbb{I}_{jt}(\theta)\, x_{it} x_{lt}\right]^2\right\} \leq M$;

(c) $\mathrm{E}\left[N^{-1}\sum_{i=1}^{N}\left\|T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{jt}(\theta)\,e_{it}\boldsymbol{\lambda}_{ji}^{0}\right\|^{2}\right]\leq M.$

Lemma A.2 *Given $\hat{\mathbf{H}}_{jj}(\theta)$ and $\hat{\mathbf{H}}_{mj}(\theta)$ defined in* (9) *and* (10)*, respectively, for $j=1,2$, and $j\neq m$, and for any θ,*

$$S_{\mathbf{F\Lambda}}(\theta)-S_{\mathbf{F}}\left[\boldsymbol{\Lambda}_1^0\hat{\mathbf{H}}_{11}(\theta)+\boldsymbol{\Lambda}_2^0\hat{\mathbf{H}}_{21}(\theta),\boldsymbol{\Lambda}_2^0\hat{\mathbf{H}}_{22}(\theta)+\boldsymbol{\Lambda}_1^0\hat{\mathbf{H}}_{12}(\theta),\theta\right]=O_p\left(C_{NT}^{-1}\right).$$

Lemma A.3 *There exists a $\tau(\theta)>0$ such that*

$$\operatorname{p}\lim\inf_{N,T\to\infty}S_{\mathbf{F}}\left[\boldsymbol{\Lambda}_1^0\hat{\mathbf{H}}_{11}(\theta)+\boldsymbol{\Lambda}_2^0\hat{\mathbf{H}}_{21}(\theta),\boldsymbol{\Lambda}_2^0\hat{\mathbf{H}}_{22}(\theta)+\boldsymbol{\Lambda}_1^0\hat{\mathbf{H}}_{12}(\theta),\theta\right]-S_{\mathbf{F}}\left(\boldsymbol{\Lambda}_1^0,\boldsymbol{\Lambda}_2^0,\theta^0\right)=\tau(\theta),\quad\forall\theta\neq\theta^0.$$

Proof of Theorem 3.1. As defined in Section 3.2, $\tilde{\mathbf{V}}_1$ is the $R^0\times R^0$ diagonal matrix of the first R^0 largest eigenvalues of $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}=(NT)^{-1}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'$ in decreasing order, and $\tilde{\boldsymbol{\Lambda}}_1$ is the estimator for $\boldsymbol{\Lambda}_1^0$ in the true data generating process $\mathbf{x}_t=\boldsymbol{\Lambda}_1^0\mathbf{f}_t^0+\mathbb{I}_{2t}\left(\theta^0\right)\boldsymbol{\Delta}^0\mathbf{f}_t^0+\mathbf{e}_t$ from the misspecified linear model $\mathbf{x}_t=\boldsymbol{\Lambda}_1\mathbf{f}_t+\mathbf{e}_t$: the equality $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}\tilde{\boldsymbol{\Lambda}}_1=\tilde{\boldsymbol{\Lambda}}_1\tilde{\mathbf{V}}_1$ then holds by the definitions of eigenvectors and eigenvalues. Applying the normalization $N^{-1}\tilde{\boldsymbol{\Lambda}}_1'\tilde{\boldsymbol{\Lambda}}_1=\mathbf{I}_{R^0}$ to implement the principal components estimator, it follows that $N^{-1}\sum_{i=1}^{N}\left\|\tilde{\boldsymbol{\lambda}}_{1i}\right\|^2=O_p(1)$. By Lemma A.3 in Bai (2003), $\tilde{\mathbf{V}}_1\xrightarrow{p}\mathbf{V}_1$ where $\mathbf{V}_1$ is a positive definite matrix: we then focus on $\left\|\tilde{\mathbf{V}}_1\left(\tilde{\boldsymbol{\lambda}}_{1i}-\tilde{\mathbf{H}}_1'\boldsymbol{\lambda}_{1i}^0\right)\right\|^2$. Theorem 3.1 relies on the identity

$$\begin{aligned}\tilde{\mathbf{V}}_1\left(\tilde{\boldsymbol{\lambda}}_{1i}-\tilde{\mathbf{H}}_1'\boldsymbol{\lambda}_{1i}^0\right)\;=\;&N^{-1}\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\sigma_{1il}\left(\theta^0\right)+N^{-1}\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\varkappa_{1il}\left(\theta^0\right)\\&+N^{-1}\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\sigma_{2il}\left(\theta^0\right)+N^{-1}\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\varkappa_{2il}\left(\theta^0\right)\\&+N^{-1}\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\varphi_{il}+N^{-1}\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\varphi_{li}+N^{-1}\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\vartheta_{il}+N^{-1}\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\vartheta_{li}+N^{-1}\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\psi_{il},\end{aligned}$$

where

$$\begin{gathered}\varkappa_{jil}(\theta)=T^{-1}\sum_{t=1}^{T}\mathbb{I}_{jt}(\theta)\,e_{it}e_{lt}-\sigma_{jil}(\theta),\quad j=1,2,\\ \varphi_{jil}(\theta)=T^{-1}\sum_{t=1}^{T}\mathbb{I}_{jt}(\theta)\left[\mathbb{I}_{1t}\left(\theta^0\right)\boldsymbol{\lambda}_{1i}^{0\prime}\mathbf{f}_t^0+\mathbb{I}_{2t}\left(\theta^0\right)\boldsymbol{\lambda}_{2i}^{0\prime}\mathbf{f}_t^0\right]e_{lt},\quad j=1,2,\\ \varphi_{il}=\varphi_{1il}(\theta)+\varphi_{2il}(\theta)=T^{-1}\sum_{t=1}^{T}\left[\mathbb{I}_{1t}\left(\theta^0\right)\boldsymbol{\lambda}_{1i}^{0\prime}\mathbf{f}_t^0+\mathbb{I}_{2t}\left(\theta^0\right)\boldsymbol{\lambda}_{2i}^{0\prime}\mathbf{f}_t^0\right]e_{lt},\\ \varphi_{jli}(\theta)=T^{-1}\sum_{t=1}^{T}\mathbb{I}_{jt}(\theta)\left[\mathbb{I}_{1t}\left(\theta^0\right)\boldsymbol{\lambda}_{1l}^{0\prime}\mathbf{f}_t^0+\mathbb{I}_{2t}\left(\theta^0\right)\boldsymbol{\lambda}_{2l}^{0}\mathbf{f}_t^0\right]'e_{it},\quad j=1,2,\\ \varphi_{li}=\varphi_{1li}(\theta)+\varphi_{2li}(\theta)=T^{-1}\sum_{t=1}^{T}\left[\mathbb{I}_{1t}\left(\theta^0\right)\boldsymbol{\lambda}_{1l}^{0\prime}\mathbf{f}_t^0+\mathbb{I}_{2t}\left(\theta^0\right)\boldsymbol{\lambda}_{2l}^{0}\mathbf{f}_t^0\right]'e_{it},\\ \vartheta_{il}=T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\theta^0\right)\boldsymbol{\lambda}_{1i}^{0\prime}\mathbf{f}_t^0\mathbf{f}_t^{0\prime}\boldsymbol{\delta}_l^0,\quad\vartheta_{li}=T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\theta^0\right)\boldsymbol{\lambda}_{1l}^{0\prime}\mathbf{f}_t^0\mathbf{f}_t^{0\prime}\boldsymbol{\delta}_i^0,\\ \psi_{il}=T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\theta^0\right)\boldsymbol{\delta}_i^{0\prime}\mathbf{f}_t^0\mathbf{f}_t^{0\prime}\boldsymbol{\delta}_l^0.\end{gathered}\tag{21}$$

The matrix $\tilde{\mathbf{H}}_1$ depends on N and T: this dependence is implicitly suppressed to keep notation simple. Notice that

$$\left\|\tilde{\mathbf{H}}_1\right\|\leq\left\|\frac{\mathbf{F}^0\mathbf{F}^{0\prime}}{T}\right\|\left\|\frac{\boldsymbol{\Lambda}_1^{0\prime}\boldsymbol{\Lambda}_1^0}{N}\right\|^{1/2}\left\|\frac{\tilde{\boldsymbol{\Lambda}}_1'\tilde{\boldsymbol{\Lambda}}_1}{N}\right\|^{1/2}\left\|\tilde{\mathbf{V}}_1^{-1}\right\|=O_p(1),$$

by Assumptions C1 and C2. By Loève's inequality,

$$N^{-1}\sum_{i=1}^{N}\left\|\tilde{\mathbf{V}}_1\left(\tilde{\boldsymbol{\lambda}}_{1i}-\tilde{\mathbf{H}}_1'\boldsymbol{\lambda}_{1i}^0\right)\right\|^2\leq 9N^{-1}\sum_{i=1}^{N}\left[\tilde{\sigma}_{1i\cdot}(\theta)+\tilde{\varkappa}_{1i\cdot}(\theta)+\tilde{\sigma}_{2i\cdot}(\theta)+\tilde{\varkappa}_{2i\cdot}(\theta)+\tilde{\varphi}_{i\cdot}+\tilde{\varphi}_{\cdot i}+\tilde{\vartheta}_{i\cdot}+\tilde{\vartheta}_{\cdot i}+\tilde{\psi}_{i\cdot}\right],$$

where

$$\tilde{\sigma}_{ji\cdot}(\theta) = N^{-2}\left\|\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\sigma_{jil}(\theta)\right\|^2, \quad \tilde{\varkappa}_{ji\cdot}(\theta) = N^{-2}\left\|\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\varkappa_{jil}(\theta)\right\|^2, \quad j=1,2,$$

$$\tilde{\varphi}_{i\cdot} = N^{-2}\left\|\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\varphi_{il}\right\|^2, \quad \tilde{\varphi}_{\cdot i} = N^{-2}\left\|\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\varphi_{li}\right\|^2,$$

$$\tilde{\vartheta}_{i\cdot} = N^{-2}\left\|\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\vartheta_{il}\right\|^2, \quad \tilde{\vartheta}_{\cdot i} = N^{-2}\left\|\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\vartheta_{li}\right\|^2,$$

$$\tilde{\psi}_{i\cdot} = N^{-2}\left\|\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\psi_{il}\right\|^2.$$

We first consider $\tilde{\sigma}_{1i\cdot}(\theta)$: $\tilde{\sigma}_{2i\cdot}(\theta)$ is analogous and omitted. We have

$$\left\|\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\sigma_{1il}(\theta)\right\|^2 \leq \left(\sum_{l=1}^{N}\left\|\tilde{\boldsymbol{\lambda}}_{1l}\right\|^2\right)\left[\sum_{l=1}^{N}\sigma_{1il}^2(\theta)\right]$$

and

$$N^{-1}\sum_{i=1}^{N}\tilde{\sigma}_{1i\cdot}(\theta) \leq N^{-1}\left(N^{-1}\sum_{l=1}^{N}\left\|\tilde{\boldsymbol{\lambda}}_{1l}\right\|^2\right)\left[N^{-1}\sum_{i=1}^{N}\sum_{l=1}^{N}\sigma_{1il}^2(\theta)\right] = O_p\left(N^{-1}\right)$$

by Lemma A.1(a). As for $\tilde{\varkappa}_{ji\cdot}(\theta)$, for $j=1$ ($j=2$ is analogous),

$$\begin{aligned}
\sum_{i=1}^{N}\tilde{\varkappa}_{1i\cdot}(\theta) &= N^{-2}\sum_{i=1}^{N}\left\|\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\varkappa_{1il}(\theta)\right\|^2 \\
&= N^{-2}\sum_{i=1}^{N}\sum_{l=1}^{N}\sum_{q=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}'\tilde{\boldsymbol{\lambda}}_{1q}\varkappa_{1il}(\theta)\varkappa_{1iq}(\theta) \\
&\leq \left[N^{-2}\sum_{l=1}^{N}\sum_{q=1}^{N}\left(\tilde{\boldsymbol{\lambda}}_{1l}'\tilde{\boldsymbol{\lambda}}_{1q}\right)^2\right]^{1/2}\left\{N^{-2}\sum_{l=1}^{N}\sum_{q=1}^{N}\left[\sum_{i=1}^{N}\varkappa_{1il}(\theta)\varkappa_{1iq}(\theta)\right]^2\right\}^{1/2} \\
&\leq \left(N^{-1}\sum_{l=1}^{N}\left\|\tilde{\boldsymbol{\lambda}}_{1l}\right\|^2\right)\left\{N^{-2}\sum_{l=1}^{N}\sum_{q=1}^{N}\left[\sum_{i=1}^{N}\varkappa_{1il}(\theta)\varkappa_{1iq}(\theta)\right]^2\right\}^{1/2}.
\end{aligned}$$

Since

$$\mathrm{E}\left\{\left[\sum_{i=1}^{N}\varkappa_{1il}(\theta)\varkappa_{1iq}(\theta)\right]^2\right\} = \mathrm{E}\left[\sum_{i=1}^{N}\sum_{u=1}^{N}\varkappa_{1il}(\theta)\varkappa_{1iq}(\theta)\varkappa_{1ul}(\theta)\varkappa_{1uq}(\theta)\right]^2 \leq N^2\max_{i,l}\mathrm{E}\left|\varkappa_{1il}(\theta)\right|^4$$

and

$$\mathrm{E}\left|\varkappa_{1il}(\theta)\right|^4 = T^{-2}\mathrm{E}\left|T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)e_{it}e_{lt} - \mathrm{E}\left[\mathbb{I}_{1t}(\theta)e_{it}e_{lt}\right]\right|^4 \leq T^{-2}M$$

by Assumption C3(d), then

$$\sum_{i=1}^{N}\tilde{\varkappa}_{1i\cdot}(\theta) \leq O_p(1)\sqrt{\frac{N^2}{T^2}} = O_p\left(\frac{N}{T}\right)$$

and $N^{-1}\sum_{i=1}^{N}\tilde{\varkappa}_{1i\cdot}(\theta) = O_p\left(T^{-1}\right)$. Regarding $\tilde{\varphi}_{i\cdot}$, we have

$$\begin{aligned}
\tilde{\varphi}_{i\cdot} &= N^{-2}\left\|\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\varphi_{il}\right\|^2 \\
&= N^{-2}\left\|\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\left\{T^{-1}\sum_{t=1}^{T}\left[\mathbb{I}_{1t}\left(\theta^0\right)\boldsymbol{\lambda}_{1i}^{0\prime}\mathbf{f}_t^0 + \mathbb{I}_{2t}\left(\theta^0\right)\boldsymbol{\lambda}_{2i}^{0\prime}\mathbf{f}_t^0\right]e_{lt}\right\}\right\|^2 \\
&\leq \left\{N^{-2}\sum_{l=1}^{N}\left\|\tilde{\boldsymbol{\lambda}}_{1l}\boldsymbol{\lambda}_{1i}^{0\prime}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\theta^0\right)\mathbf{f}_t^0e_{lt}\right]\right\|^2\right\} + \left\{N^{-2}\sum_{l=1}^{N}\left\|\tilde{\boldsymbol{\lambda}}_{1l}\boldsymbol{\lambda}_{2i}^{0\prime}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\theta^0\right)\mathbf{f}_t^0e_{lt}\right]\right\|^2\right\} \\
&\leq \left\{N^{-1}\sum_{l=1}^{N}\left[T^{-2}\left\|\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\theta^0\right)\mathbf{f}_t^0e_{lt}\right\|^2\right]\right\}\left\|\boldsymbol{\lambda}_{1i}^0\right\|^2\left(N^{-1}\sum_{l=1}^{N}\left\|\tilde{\boldsymbol{\lambda}}_{1l}\right\|^2\right) \\
&\quad + \left\{N^{-1}\sum_{l=1}^{N}\left[T^{-2}\left\|\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\theta^0\right)\mathbf{f}_t^0e_{lt}\right\|^2\right]\right\}\left\|\boldsymbol{\lambda}_{2i}^0\right\|^2\left(N^{-1}\sum_{l=1}^{N}\left\|\tilde{\boldsymbol{\lambda}}_{1l}\right\|^2\right) \\
&\leq \left\{\left\{N^{-1}\sum_{l=1}^{N}\left[T^{-2}\left\|\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\theta^0\right)\mathbf{f}_t^0e_{lt}\right\|^2\right]\right\}\left\|\boldsymbol{\lambda}_{1i}^0\right\|^2 + \left\{N^{-1}\sum_{l=1}^{N}\left[T^{-2}\left\|\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\theta^0\right)\mathbf{f}_t^0e_{lt}\right\|^2\right]\right\}\left\|\boldsymbol{\lambda}_{2i}^0\right\|^2\right\}O_p(1)
\end{aligned}$$

and

$$N^{-1}\sum_{i=1}^{N}\tilde{\varphi}_{i\cdot}=\left\{\left\{N^{-1}\sum_{l=1}^{N}\left[T^{-2}\left\|\sum_{t=1}^{T}\mathbb{I}_{1t}\left(\theta^{0}\right)\mathbf{f}_{t}^{0}e_{lt}\right\|^{2}\right]\right\}O_{p}\left(1\right)+\left\{N^{-1}\sum_{l=1}^{N}\left[T^{-2}\left\|\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\theta^{0}\right)\mathbf{f}_{t}^{0}e_{lt}\right\|^{2}\right]\right\}O_{p}\left(1\right)\right\}O_{p}\left(1\right)=O_{p}\left(T^{-1}\right)$$

by Assumptions C2 and C4. In a similar way, it is proved that $\tilde{\varphi}_{\cdot i}=O_p\left(T^{-1}\right)$. As for $\tilde{\vartheta}_{i\cdot}$,

$$\begin{aligned}
\tilde{\vartheta}_{i\cdot} &= N^{-2}\left\|\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\vartheta_{il}\right\|^{2}\\
&= N^{-2}\left(\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\vartheta_{il}\right)'\left(\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\vartheta_{il}\right)\\
&= N^{-1}\left\{\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\boldsymbol{\lambda}_{1i}^{0\prime}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\theta^{0}\right)\mathbf{f}_{t}^{0}\mathbf{f}_{t}^{0\prime}\boldsymbol{\delta}_{l}^{0}\right]\right\}'N^{-1}\left\{\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\boldsymbol{\lambda}_{1i}^{0\prime}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\theta^{0}\right)\mathbf{f}_{t}^{0}\mathbf{f}_{t}^{0\prime}\boldsymbol{\delta}_{l}^{0}\right]\right\}\\
&= N^{-1}\left\{\sum_{l=1}^{N^{\alpha^0}}\tilde{\boldsymbol{\lambda}}_{1l}\boldsymbol{\lambda}_{1i}^{0\prime}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\theta^{0}\right)\mathbf{f}_{t}^{0}\mathbf{f}_{t}^{0\prime}\boldsymbol{\delta}_{l}^{0}\right]+\sum_{l=N^{\alpha^0}+1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\boldsymbol{\lambda}_{1i}^{0\prime}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\theta^{0}\right)\mathbf{f}_{t}^{0}\mathbf{f}_{t}^{0\prime}\boldsymbol{\delta}_{l}^{0}\right]\right\}'\\
&\quad\times N^{-1}\left\{\sum_{l=1}^{N^{\alpha^0}}\tilde{\boldsymbol{\lambda}}_{1l}\boldsymbol{\lambda}_{1i}^{0\prime}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\theta^{0}\right)\mathbf{f}_{t}^{0}\mathbf{f}_{t}^{0\prime}\boldsymbol{\delta}_{l}^{0}\right]+\sum_{l=N^{\alpha^0}+1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\boldsymbol{\lambda}_{1i}^{0\prime}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\theta^{0}\right)\mathbf{f}_{t}^{0}\mathbf{f}_{t}^{0\prime}\boldsymbol{\delta}_{l}^{0}\right]\right\}\\
&= N^{-1}O_p\left(N^{\alpha^0}\right)\cdot N^{-1}O_p\left(N^{\alpha^0}\right)\\
&= O_p\left(N^{2\alpha^0-2}\right).
\end{aligned}$$

In a similar way, it can be proved that $\tilde{\vartheta}_{\cdot i}=O_p\left(N^{2\alpha^0-2}\right)$. Finally, under Assumptions C1 and C2,

$$\begin{aligned}
\tilde{\psi}_{i\cdot} &= N^{-2}\left\|\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\psi_{il}\right\|^{2}\\
&= N^{-2}\left(\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\psi_{il}\right)'\left(\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\psi_{il}\right)\\
&= N^{-1}\left\{\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\theta^{0}\right)\boldsymbol{\delta}_{i}^{0\prime}\mathbf{f}_{t}^{0}\mathbf{f}_{t}^{0\prime}\boldsymbol{\delta}_{l}^{0}\right]\right\}'N^{-1}\left\{\sum_{l=1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\theta^{0}\right)\boldsymbol{\delta}_{i}^{0\prime}\mathbf{f}_{t}^{0}\mathbf{f}_{t}^{0\prime}\boldsymbol{\delta}_{l}^{0}\right]\right\}\\
&= N^{-1}\left\{\sum_{l=1}^{N^{\alpha^0}}\tilde{\boldsymbol{\lambda}}_{1l}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\theta^{0}\right)\boldsymbol{\delta}_{i}^{0\prime}\mathbf{f}_{t}^{0}\mathbf{f}_{t}^{0\prime}\boldsymbol{\delta}_{l}^{0}\right]+\sum_{l=N^{\alpha^0}+1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\theta^{0}\right)\boldsymbol{\delta}_{i}^{0\prime}\mathbf{f}_{t}^{0}\mathbf{f}_{t}^{0\prime}\boldsymbol{\delta}_{l}^{0}\right]\right\}'\\
&\quad\times N^{-1}\left\{\sum_{l=1}^{N^{\alpha^0}}\tilde{\boldsymbol{\lambda}}_{1l}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\theta^{0}\right)\boldsymbol{\delta}_{i}^{0\prime}\mathbf{f}_{t}^{0}\mathbf{f}_{t}^{0\prime}\boldsymbol{\delta}_{l}^{0}\right]+\sum_{l=N^{\alpha^0}+1}^{N}\tilde{\boldsymbol{\lambda}}_{1l}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}\left(\theta^{0}\right)\boldsymbol{\delta}_{i}^{0\prime}\mathbf{f}_{t}^{0}\mathbf{f}_{t}^{0\prime}\boldsymbol{\delta}_{l}^{0}\right]\right\}\\
&= N^{-1}O_p\left(N^{\alpha^0}\right)N^{-1}O_p\left(N^{\alpha^0}\right)\\
&= O_p\left(N^{2\alpha^0-2}\right).
\end{aligned}$$

Combining all above results, we have

$$N^{-1}\sum_{i=1}^{N}\left\|\tilde{\mathbf{V}}_{1}\left(\tilde{\boldsymbol{\lambda}}_{1i}-\tilde{\mathbf{H}}_{1}^{\prime}\boldsymbol{\lambda}_{1i}^{0}\right)\right\|^{2}=O_{p}\left(N^{-1}\right)+O_{p}\left(T^{-1}\right)+O_{p}\left(N^{2\alpha^{0}-2}\right),$$

which completes the proof of the theorem. ■

Proof of Theorem 3.2. From Theorem 3.1, by Assumption I the regime indicator $\mathbb{I}_{jt}\left(\theta\right)$ is identified, for $j=1,2$: we can then split the sample according to the value of $\mathbb{I}_{jt}\left(\theta\right)$. We consider the case $j=1$: the case $j=2$ is analogous and omitted. As defined in Section 3.4, $\hat{\mathbf{V}}_1\left(\theta\right)$ is the $R^0\times R^0$ diagonal matrix of the first R^0 largest eigenvalues of $\hat{\boldsymbol{\Sigma}}_{1\mathbf{x}}\left(\theta\right)$ in (7) in decreasing order: the equality $\hat{\boldsymbol{\Sigma}}_{1\mathbf{x}}\left(\theta\right)\hat{\boldsymbol{\Lambda}}_1\left(\theta\right)=\hat{\boldsymbol{\Lambda}}_1\left(\theta\right)\hat{\mathbf{V}}_1\left(\theta\right)$ holds by the definitions of eigenvectors and eigenvalues. From the normalization $N^{-1}\hat{\boldsymbol{\Lambda}}_1\left(\theta\right)'\hat{\boldsymbol{\Lambda}}_1\left(\theta\right)=\mathbf{I}_{R^0}$, it follows that $N^{-1}\sum_{i=1}^{N}\left\|\hat{\boldsymbol{\lambda}}_{1i}\left(\theta\right)\right\|^2=O_p\left(1\right)$ for all θ. By Lemma A.3 in Bai (2003), $\hat{\mathbf{V}}_1\left(\theta\right)\xrightarrow{p}\mathbf{V}_1\left(\theta\right)$ where $\mathbf{V}_1\left(\theta\right)$ is a positive definite matrix for all θ, and $\left\|\hat{\mathbf{V}}_1\left(\theta\right)\right\|=O_p\left(1\right)$: we then focus on $\left\|\hat{\mathbf{V}}_1(\theta)\left[\hat{\boldsymbol{\lambda}}_{1i}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right]\right\|^2$. Theorem 3.2 relies on the identity

$$\begin{aligned}\hat{\mathbf{V}}_1(\theta)\left[\hat{\boldsymbol{\lambda}}_{1i}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right] &= N^{-1}\sum_{l=1}^N\hat{\boldsymbol{\lambda}}_{1l}(\theta)\sigma_{1il}(\theta)+N^{-1}\sum_{l=1}^N\hat{\boldsymbol{\lambda}}_{1l}(\theta)\varkappa_{1il}(\theta)\\ &\quad+N^{-1}\sum_{l=1}^N\hat{\boldsymbol{\lambda}}_{1l}(\theta)\varphi_{1il}(\theta)+N^{-1}\sum_{l=1}^N\hat{\boldsymbol{\lambda}}_{1l}(\theta)\varphi_{1li}(\theta)\end{aligned},$$

where $\varkappa_{1il}(\theta)$, $\varphi_{1il}(\theta)$ and $\varphi_{1li}(\theta)$ are defined in (21). The matrices $\hat{\mathbf{H}}_{11}(\theta)$ and $\hat{\mathbf{H}}_{21}(\theta)$ both depend on N and T: this dependence is implicitly suppressed to keep notation simple. Notice that

$$\left\|\hat{\mathbf{H}}_{11}(\theta)\right\|\le\left\|\frac{\mathbf{F}_1^0(\theta^0)\mathbf{F}_1^0(\theta)'}{T}\right\|\left\|\frac{\boldsymbol{\Lambda}_1^{0\prime}\boldsymbol{\Lambda}_1^0}{N}\right\|^{1/2}\left\|\frac{\hat{\boldsymbol{\Lambda}}_1(\theta)'\hat{\boldsymbol{\Lambda}}_1(\theta)}{N}\right\|^{1/2}\left\|\hat{\mathbf{V}}_1(\theta)^{-1}\right\|=O_p(1)$$

by Assumptions C1 and C2. In an analogous way, it can be shown that $\left\|\hat{\mathbf{H}}_{21}(\theta)\right\|=O_p(1)$. By Loève's inequality

$$N^{-1}\sum_{i=1}^N\left\|\hat{\mathbf{V}}_1(\theta)\left[\hat{\boldsymbol{\lambda}}_{1i}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right]\right\|^2\le 4N^{-1}\sum_{i=1}^N\left[\hat{\sigma}_{1i\cdot}(\theta)+\hat{\varkappa}_{1i\cdot}(\theta)+\hat{\varphi}_{1i\cdot}(\theta)+\hat{\varphi}_{1\cdot i}(\theta)\right],$$

where

$$\begin{aligned}\hat{\sigma}_{1i\cdot}(\theta)&=N^{-2}\left\|\sum_{l=1}^N\hat{\boldsymbol{\lambda}}_{1l}(\theta)\sigma_{1il}(\theta)\right\|^2, &\hat{\varkappa}_{1i\cdot}(\theta)&=N^{-2}\left\|\sum_{l=1}^N\hat{\boldsymbol{\lambda}}_{1l}(\theta)\varkappa_{1il}(\theta)\right\|^2,\\ \hat{\varphi}_{1i\cdot}(\theta)&=N^{-2}\left\|\sum_{l=1}^N\hat{\boldsymbol{\lambda}}_{1l}(\theta)\varphi_{1il}(\theta)\right\|^2, &\hat{\varphi}_{1\cdot i}(\theta)&=N^{-2}\left\|\sum_{l=1}^N\hat{\boldsymbol{\lambda}}_{1l}(\theta)\varphi_{1li}(\theta)\right\|^2.\end{aligned}$$

Starting from $\hat{\sigma}_{1i\cdot}(\theta)$,

$$\left\|\sum_{l=1}^N\hat{\boldsymbol{\lambda}}_{1l}(\theta)\sigma_{1il}(\theta)\right\|^2\le\left[\sum_{l=1}^N\left\|\hat{\boldsymbol{\lambda}}_{1l}(\theta)\right\|^2\right]\left[\sum_{l=1}^N\sigma_{1il}^2(\theta)\right],$$

and

$$N^{-1}\sum_{i=1}^N\hat{\sigma}_{1i\cdot}(\theta)\le N^{-1}\left[N^{-1}\sum_{l=1}^N\left\|\hat{\boldsymbol{\lambda}}_{1l}(\theta)\right\|^2\right]\left[N^{-1}\sum_{i=1}^N\sum_{l=1}^N\sigma_{1il}^2(\theta)\right]=O_p\left(N^{-1}\right)$$

by Lemma A.1(a). As for $\hat{\varkappa}_{1i\cdot}(\theta)$,

$$\begin{aligned}\sum_{i=1}^N\hat{\varkappa}_{1i\cdot}(\theta)&=N^{-2}\sum_{i=1}^N\left\|\sum_{l=1}^N\hat{\boldsymbol{\lambda}}_{1l}(\theta)\varkappa_{1il}(\theta)\right\|^2\\ &=N^{-2}\sum_{i=1}^N\sum_{l=1}^N\sum_{q=1}^N\hat{\boldsymbol{\lambda}}_{1l}(\theta)'\hat{\boldsymbol{\lambda}}_{1q}(\theta)\varkappa_{1il}(\theta)\varkappa_{1iq}(\theta)\\ &\le\left\{N^{-2}\sum_{l=1}^N\sum_{q=1}^N\left[\hat{\boldsymbol{\lambda}}_{1l}(\theta)'\hat{\boldsymbol{\lambda}}_{1q}(\theta)\right]^2\right\}^{1/2}\left\{N^{-2}\sum_{l=1}^N\sum_{q=1}^N\left[\sum_{i=1}^N\varkappa_{1il}(\theta)\varkappa_{1iq}(\theta)\right]^2\right\}^{1/2}\\ &\le\left[N^{-1}\sum_{l=1}^N\left\|\hat{\boldsymbol{\lambda}}_{1l}(\theta)\right\|^2\right]\left\{N^{-2}\sum_{l=1}^N\sum_{q=1}^N\left[\sum_{i=1}^N\varkappa_{1il}(\theta)\varkappa_{1iq}(\theta)\right]^2\right\}^{1/2}.\end{aligned}$$

Since

$$\mathrm{E}\left\{\left[\sum_{i=1}^N\varkappa_{1il}(\theta)\varkappa_{1iq}(\theta)\right]^2\right\}=\mathrm{E}\left[\sum_{i=1}^N\sum_{u=1}^N\varkappa_{1il}(\theta)\varkappa_{1iq}(\theta)\varkappa_{1ul}(\theta)\varkappa_{1uq}(\theta)\right]\le N^2\max_{i,l}\mathrm{E}\left|\varkappa_{1il}(\theta)\right|^4$$

and

$$\mathrm{E}\left|\varkappa_{1il}(\theta)\right|^4=T^{-2}\mathrm{E}\left|T^{-1/2}\sum_{t=1}^T\mathbb{I}_{1t}(\theta)e_{it}e_{lt}-\mathrm{E}\left[\mathbb{I}_{1t}(\theta)e_{it}e_{lt}\right]\right|^4\le T^{-2}M$$

by Assumption C3(d), then

$$\sum_{i=1}^N\hat{\varkappa}_{1i\cdot}(\theta)\le O_p(1)\sqrt{\frac{N^2}{T^2}}=O_p\left(\frac{N}{T}\right)$$

and $N^{-1}\sum_{i=1}^{N}\hat{\varkappa}_{1i\cdot}(\theta)=O_p\left(T^{-1}\right)$. Regarding $\hat{\varphi}_{1i\cdot}(\theta)$,

$$\begin{aligned}
\hat{\varphi}_{1i\cdot}(\theta) &= N^{-2}\left\|\sum_{l=1}^{N}\hat{\boldsymbol{\lambda}}_{1l}(\theta)\varphi_{1il}(\theta)\right\|^2 \\
&= N^{-2}\left\|\sum_{l=1}^{N}\hat{\boldsymbol{\lambda}}_{1l}(\theta)\left\{T^{-1}\sum_{t=1}^{T}\left\{\mathbb{I}_{1t}(\theta)\left[\mathbb{I}_{1t}\left(\theta^0\right)\boldsymbol{\lambda}_{1i}^{0\prime}\mathbf{f}_t^0+\mathbb{I}_{2t}\left(\theta^0\right)\boldsymbol{\lambda}_{2i}^{0\prime}\mathbf{f}_t^0\right]e_{lt}\right\}\right\}\right\|^2 \\
&\leq N^{-2}\sum_{l=1}^{N}\left\|\hat{\boldsymbol{\lambda}}_{1l}(\theta)\boldsymbol{\lambda}_{1i}^{0\prime}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)\mathbb{I}_{1t}\left(\theta^0\right)\mathbf{f}_t^0e_{lt}\right]\right\|^2+N^{-2}\sum_{l=1}^{N}\left\|\hat{\boldsymbol{\lambda}}_{1l}(\theta)\boldsymbol{\lambda}_{2i}^{0\prime}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)\mathbb{I}_{2t}\left(\theta^0\right)\mathbf{f}_t^0e_{lt}\right]\right\|^2 \\
&\leq \left\{N^{-1}\sum_{l=1}^{N}\left[T^{-2}\left\|\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)\mathbf{f}_t^0e_{lt}\right\|^2\right]\right\}\left\|\boldsymbol{\lambda}_{1i}^0\right\|^2\left(N^{-1}\sum_{l=1}^{N}\left\|\hat{\boldsymbol{\lambda}}_{1l}(\theta)\right\|^2\right) \\
&\quad+\left\{N^{-1}\sum_{l=1}^{N}\left[T^{-2}\left\|\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)\mathbf{f}_t^0e_{lt}\right\|^2\right]\right\}\left\|\boldsymbol{\lambda}_{2i}^0\right\|^2\left(N^{-1}\sum_{l=1}^{N}\left\|\hat{\boldsymbol{\lambda}}_{1l}(\theta)\right\|^2\right) \\
&\leq \left\{\left\{N^{-1}\sum_{l=1}^{N}\left[T^{-2}\left\|\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)\mathbf{f}_t^0e_{lt}\right\|^2\right]\right\}\left\|\boldsymbol{\lambda}_{1i}^0\right\|^2+\left\{N^{-1}\sum_{l=1}^{N}\left[T^{-2}\left\|\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)\mathbf{f}_t^0e_{lt}\right\|^2\right]\right\}\left\|\boldsymbol{\lambda}_{2i}^0\right\|^2\right\}O_p(1)
\end{aligned}$$

and

$$N^{-1}\sum_{i=1}^{N}\hat{\varphi}_{1i\cdot}(\theta)=\left\{\left\{N^{-1}\sum_{l=1}^{N}\left[T^{-2}\left\|\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)\mathbf{f}_t^0e_{lt}\right\|^2\right]\right\}O_p(1)+\left\{N^{-1}\sum_{l=1}^{N}\left[T^{-2}\left\|\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)\mathbf{f}_t^0e_{lt}\right\|^2\right]\right\}O_p(1)\right\}O_p(1)=O_p\left(T^{-1}\right)$$

by Assumptions C2 and C4. In an analogous way, it can be proved that

$$N^{-1}\sum_{i=1}^{N}\hat{\varphi}_{1\cdot i}(\theta)=O_p\left(T^{-1}\right).$$

Combining all results above, we have

$$N^{-1}\sum_{i=1}^{N}\left\|\hat{\mathbf{V}}_1(\theta)\left[\hat{\boldsymbol{\lambda}}_{1i}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right]\right\|^2=O_p\left(N^{-1}\right)+O_p\left(T^{-1}\right).$$

This completes the proof of the theorem. ∎

Proof of Theorem 3.3. In order to prove the theorem, it is sufficient to prove that

$$\lim_{N,T\to\infty}\mathrm{P}\left[S_{\mathbf{F}\boldsymbol{\Lambda}}(\theta)\leq S_{\mathbf{F}\boldsymbol{\Lambda}}\left(\theta^0\right)\right]=0,\quad\forall\theta\neq\theta^0,$$

where $S_{\mathbf{F}\boldsymbol{\Lambda}}(\theta)$ is defined in (8). Consider the identity

$$\begin{aligned}
S_{\mathbf{F}\boldsymbol{\Lambda}}(\theta)-S_{\mathbf{F}\boldsymbol{\Lambda}}\left(\theta^0\right) &= S_{\mathbf{F}\boldsymbol{\Lambda}}(\theta)-S_{\mathbf{F}}\left[\boldsymbol{\Lambda}_1^0\hat{\mathbf{H}}_{11}(\theta)+\boldsymbol{\Lambda}_2^0\hat{\mathbf{H}}_{21}(\theta),\boldsymbol{\Lambda}_2^0\hat{\mathbf{H}}_{22}(\theta)+\boldsymbol{\Lambda}_1^0\hat{\mathbf{H}}_{12}(\theta),\theta\right] \\
&\quad+S_{\mathbf{F}}\left[\boldsymbol{\Lambda}_1^0\hat{\mathbf{H}}_{11}(\theta)+\boldsymbol{\Lambda}_2^0\hat{\mathbf{H}}_{21}(\theta),\boldsymbol{\Lambda}_2^0\hat{\mathbf{H}}_{22}(\theta)+\boldsymbol{\Lambda}_1^0\hat{\mathbf{H}}_{12}(\theta),\theta\right]-S_{\mathbf{F}}\left[\boldsymbol{\Lambda}_1^0\hat{\mathbf{H}}_{11}\left(\theta^0\right),\boldsymbol{\Lambda}_2^0\hat{\mathbf{H}}_{22}\left(\theta^0\right),\theta^0\right] \\
&\quad+S_{\mathbf{F}}\left[\boldsymbol{\Lambda}_1^0\hat{\mathbf{H}}_{11}\left(\theta^0\right),\boldsymbol{\Lambda}_2^0\hat{\mathbf{H}}_{22}\left(\theta^0\right),\theta^0\right]-S_{\mathbf{F}\boldsymbol{\Lambda}}\left(\theta^0\right),
\end{aligned}$$

where $S_{\mathbf{F}}(\boldsymbol{\Lambda},\theta)$ is defined in (4). By Lemma A.2, $S_{\mathbf{F}\boldsymbol{\Lambda}}(\theta)-S_{\mathbf{F}}\left[\boldsymbol{\Lambda}_1^0\hat{\mathbf{H}}_{11}(\theta)+\boldsymbol{\Lambda}_2^0\hat{\mathbf{H}}_{21}(\theta),\boldsymbol{\Lambda}_2^0\hat{\mathbf{H}}_{22}(\theta)+\boldsymbol{\Lambda}_1^0\hat{\mathbf{H}}_{12}(\theta),\theta\right]=O_p\left(C_{NT}^{-1}\right)$ for any θ, including $\theta=\theta^0$. Since $\boldsymbol{\Lambda}_1^0\hat{\mathbf{H}}_{11}\left(\theta^0\right)$ and $\boldsymbol{\Lambda}_2^0\hat{\mathbf{H}}_{22}\left(\theta^0\right)$ span the same column space as $\boldsymbol{\Lambda}_1^0$ and $\boldsymbol{\Lambda}_2^0$, respectively, we have

$$S_{\mathbf{F}}\left[\boldsymbol{\Lambda}_1^0\hat{\mathbf{H}}_{11}\left(\theta^0\right),\boldsymbol{\Lambda}_2^0\hat{\mathbf{H}}_{22}\left(\theta^0\right),\theta^0\right]=S_{\mathbf{F}}\left(\boldsymbol{\Lambda}_1^0,\boldsymbol{\Lambda}_2^0,\theta^0\right)$$

and $S_{\mathbf{F}}\left[\boldsymbol{\Lambda}_1^0\hat{\mathbf{H}}_{11}(\theta)+\boldsymbol{\Lambda}_2^0\hat{\mathbf{H}}_{21}(\theta),\boldsymbol{\Lambda}_2^0\hat{\mathbf{H}}_{22}(\theta)+\boldsymbol{\Lambda}_1^0\hat{\mathbf{H}}_{12}(\theta),\theta\right]-S_{\mathbf{F}}\left(\boldsymbol{\Lambda}_1^0,\boldsymbol{\Lambda}_2^0,\theta^0\right)$ has a positive limit by Lemma A.3. This completes the proof of the theorem. ∎

Proof of Corollary 3.1. Corollary 3.1 easily follows from Theorem 3.3 and the proof is omitted. ∎

Proof of Lemma A.1. Consider $j=1$ ($j=2$ is analogous and omitted). As for (a), let $\rho_{1il}(\theta)=\sigma_{1il}(\theta)\Big/[\sigma_{1ii}(\theta)\sigma_{1ll}(\theta)]^{1/2}$ such that $|\rho_{1il}(\theta)|\leq 1$: since $|\sigma_{1ll}(\theta)|\leq M$ for all l by Assumption C3(c), then

$$\begin{aligned}
N^{-1}\sum_{i=1}^{N}\sum_{l=1}^{N}\sigma_{1il}^{2}(\theta) &= N^{-1}\sum_{i=1}^{N}\sum_{l=1}^{N}\sigma_{1ii}(\theta)\sigma_{1ll}(\theta)\rho_{1il}^{2}(\theta)\\
&\leq MN^{-1}\sum_{i=1}^{N}\sum_{l=1}^{N}|\sigma_{1ii}(\theta)\sigma_{1ll}(\theta)|^{1/2}\,|\rho_{1il}(\theta)|\\
&= MN^{-1}\sum_{i=1}^{N}\sum_{l=1}^{N}|\sigma_{1il}(\theta)|\leq M^{2}
\end{aligned}$$

by Assumption C3(c). In order to prove (b), for $j=1$ (the proof for $j=2$ is analogous) it is sufficient to prove that $\mathrm{E}\,|\mathbb{I}_{1t}(\theta)x_{it}|^{4}\leq M$ for all (θ,i,t): we then have

$$\begin{aligned}
\mathrm{E}\,|\mathbb{I}_{1t}(\theta)x_{it}|^{4} &= \mathrm{E}\left|\mathbb{I}_{1t}(\theta)\left[\mathbb{I}_{1t}(\theta^{0})\boldsymbol{\lambda}_{1i}^{0\prime}\mathbf{f}_{t}^{0}+\mathbb{I}_{2t}(\theta^{0})\boldsymbol{\lambda}_{2i}^{0\prime}\mathbf{f}_{t}^{0}+e_{it}\right]\right|^{4}\\
&\leq \mathrm{E}\left\{\left[\mathbb{I}_{1t}(\theta)\mathbb{I}_{1t}(\theta^{0})\boldsymbol{\lambda}_{1i}^{0\prime}\mathbf{f}_{t}^{0}\right]^{4}\right\}+\mathrm{E}\left\{\left[\mathbb{I}_{1t}(\theta)\mathbb{I}_{2t}(\theta^{0})\boldsymbol{\lambda}_{2i}^{0\prime}\mathbf{f}_{t}^{0}\right]^{4}\right\}+\mathrm{E}\,|\mathbb{I}_{1t}(\theta)e_{it}|^{4}\\
&\leq \bar{\lambda}^{4}\mathrm{E}\left\|\mathbb{I}_{1t}(\theta)\mathbb{I}_{1t}(\theta^{0})\mathbf{f}_{t}^{0}\right\|^{4}+\bar{\lambda}^{4}\mathrm{E}\left\|\mathbb{I}_{1t}(\theta)\mathbb{I}_{2t}(\theta^{0})\mathbf{f}_{t}^{0}\right\|^{4}+\mathrm{E}\,|\mathbb{I}_{1t}(\theta)e_{it}|^{4}\\
&\leq M
\end{aligned}$$

by Assumptions C1, C2 and C3(a). As for (c), set $j=1$ (the proof for $j=2$ is analogous and omitted) and consider

$$\begin{aligned}
\mathrm{E}\left\|T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)e_{it}\boldsymbol{\lambda}_{1i}^{0}\right\|^{2} &= T^{-1}\sum_{t=1}^{T}\sum_{v=1}^{T}\mathrm{E}\left[\mathbb{I}_{1t}(\theta)\mathbb{I}_{1v}(\theta)e_{it}e_{iv}\right]\boldsymbol{\lambda}_{1i}^{0\prime}\boldsymbol{\lambda}_{1i}^{0}\\
&\leq \bar{\lambda}^{2}T^{-1}\sum_{t=1}^{T}\sum_{v=1}^{T}|\tau_{jtv}|\leq\bar{\lambda}^{2}M
\end{aligned}$$

by Assumptions C2 and C3(b). ■

Proof of Lemma A.2. Given $\hat{\boldsymbol{\Lambda}}_{j}(\theta)$ defined in (5), for $j=1,2$, define

$$\begin{aligned}
&\mathbf{P}_{\hat{\boldsymbol{\Lambda}}_{j}}(\theta)=\hat{\boldsymbol{\Lambda}}_{j}(\theta)\left[\hat{\boldsymbol{\Lambda}}_{j}(\theta)'\hat{\boldsymbol{\Lambda}}_{j}(\theta)\right]^{-1}\hat{\boldsymbol{\Lambda}}_{j}(\theta)',\\
&\mathbf{P}_{\boldsymbol{\Lambda}_{j}^{0}\hat{\mathbf{H}}_{jj}+\boldsymbol{\Lambda}_{m}^{0}\hat{\mathbf{H}}_{mj}}(\theta) = \left[\boldsymbol{\Lambda}_{j}^{0}\hat{\mathbf{H}}_{jj}(\theta)+\boldsymbol{\Lambda}_{m}^{0}\hat{\mathbf{H}}_{mj}(\theta)\right] \qquad j,m=1,2,\\
&\qquad\times\left\{\left[\boldsymbol{\Lambda}_{j}^{0}\hat{\mathbf{H}}_{jj}(\theta)+\boldsymbol{\Lambda}_{m}^{0}\hat{\mathbf{H}}_{mj}(\theta)\right]'\left[\boldsymbol{\Lambda}_{j}^{0}\hat{\mathbf{H}}_{jj}(\theta)+\boldsymbol{\Lambda}_{m}^{0}\hat{\mathbf{H}}_{mj}(\theta)\right]\right\}^{-1}, \qquad j\neq m,\\
&\qquad\times\left[\boldsymbol{\Lambda}_{j}^{0}\hat{\mathbf{H}}_{jj}(\theta)+\boldsymbol{\Lambda}_{m}^{0}\hat{\mathbf{H}}_{mj}(\theta)\right]'
\end{aligned} \tag{22}$$

so that

$$S_{\mathbf{F}\boldsymbol{\Lambda}}(\theta)=(NT)^{-1}\sum_{t=1}^{T}\mathbf{x}_{t}'\left\{\mathbf{I}_{N}-\left[\mathbb{I}_{1t}(\theta)\mathbf{P}_{\hat{\boldsymbol{\Lambda}}_{1}}(\theta)+\mathbb{I}_{2t}(\theta)\mathbf{P}_{\hat{\boldsymbol{\Lambda}}_{2}}(\theta)\right]\right\}\mathbf{x}_{t}$$

and

$$\begin{aligned}
&S_{\mathbf{F}}\left[\boldsymbol{\Lambda}_{1}^{0}\hat{\mathbf{H}}_{11}(\theta)+\boldsymbol{\Lambda}_{2}^{0}\hat{\mathbf{H}}_{21}(\theta),\boldsymbol{\Lambda}_{2}^{0}\hat{\mathbf{H}}_{22}(\theta)+\boldsymbol{\Lambda}_{1}^{0}\hat{\mathbf{H}}_{12}(\theta),\theta\right]\\
=\;&(NT)^{-1}\sum_{t=1}^{T}\mathbf{x}_{t}'\left\{\mathbf{I}_{N}-\left[\mathbb{I}_{1t}(\theta)\mathbf{P}_{\boldsymbol{\Lambda}_{1}^{0}\hat{\mathbf{H}}_{11}+\boldsymbol{\Lambda}_{2}^{0}\hat{\mathbf{H}}_{21}}(\theta)+\mathbb{I}_{2t}(\theta)\mathbf{P}_{\boldsymbol{\Lambda}_{2}^{0}\hat{\mathbf{H}}_{22}+\boldsymbol{\Lambda}_{1}^{0}\hat{\mathbf{H}}_{12}}(\theta)\right]\right\}\mathbf{x}_{t},
\end{aligned}$$

where $S_{\mathbf{F}}(\boldsymbol{\Lambda},\theta)$ and $S_{\mathbf{F}\boldsymbol{\Lambda}}(\theta)$ are defined in (4) and (8), respectively: it follows that

$$\begin{aligned}
&S_{\mathbf{F}\boldsymbol{\Lambda}}(\theta)-S_{\mathbf{F}}\left[\boldsymbol{\Lambda}_{1}^{0}\hat{\mathbf{H}}_{11}(\theta)+\boldsymbol{\Lambda}_{2}^{0}\hat{\mathbf{H}}_{21}(\theta),\boldsymbol{\Lambda}_{2}^{0}\hat{\mathbf{H}}_{22}(\theta)+\boldsymbol{\Lambda}_{1}^{0}\hat{\mathbf{H}}_{12}(\theta),\theta\right]\\
=\;&(NT)^{-1}\sum_{t=1}^{T}\mathbf{x}_{t}'\mathbb{I}_{1t}(\theta)\left[\mathbf{P}_{\boldsymbol{\Lambda}_{1}^{0}\hat{\mathbf{H}}_{11}+\boldsymbol{\Lambda}_{2}^{0}\hat{\mathbf{H}}_{21}}(\theta)-\mathbf{P}_{\hat{\boldsymbol{\Lambda}}_{1}}(\theta)\right]\mathbf{x}_{t}+(NT)^{-1}\sum_{t=1}^{T}\mathbf{x}_{t}'\mathbb{I}_{2t}(\theta)\left[\mathbf{P}_{\boldsymbol{\Lambda}_{2}^{0}\hat{\mathbf{H}}_{22}+\boldsymbol{\Lambda}_{1}^{0}\hat{\mathbf{H}}_{12}}(\theta)-\mathbf{P}_{\hat{\boldsymbol{\Lambda}}_{2}}(\theta)\right]\mathbf{x}_{t}.
\end{aligned}$$

Let

$$\begin{aligned}
&\mathbf{D}_{\hat{\boldsymbol{\Lambda}}_{j}}(\theta)=N^{-1}\hat{\boldsymbol{\Lambda}}_{j}(\theta)'\hat{\boldsymbol{\Lambda}}_{j}(\theta),\\
&\mathbf{D}_{\boldsymbol{\Lambda}_{j}^{0}\hat{\mathbf{H}}_{jj}+\boldsymbol{\Lambda}_{m}^{0}\hat{\mathbf{H}}_{mj}}(\theta)=N^{-1}\left[\boldsymbol{\Lambda}_{j}^{0}\hat{\mathbf{H}}_{jj}(\theta)+\boldsymbol{\Lambda}_{m}^{0}\hat{\mathbf{H}}_{mj}(\theta)\right]'\left[\boldsymbol{\Lambda}_{j}^{0}\hat{\mathbf{H}}_{jj}(\theta)+\boldsymbol{\Lambda}_{m}^{0}\hat{\mathbf{H}}_{mj}(\theta)\right],
\end{aligned} \qquad j=1,2,$$

so that for $j=1,2$ and $j\neq m$,

$$\begin{aligned}
\mathbf{P}_{\hat{\mathbf{\Lambda}}_j}(\theta)-\mathbf{P}_{\mathbf{\Lambda}_j^0\hat{\mathbf{H}}_{jj}+\mathbf{\Lambda}_m^0\hat{\mathbf{H}}_{mj}}(\theta) &= N^{-1}\hat{\mathbf{\Lambda}}_j(\theta)\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_j}(\theta)\right]^{-1}\hat{\mathbf{\Lambda}}_j(\theta)' \\
&\quad -N^{-1}\left[\mathbf{\Lambda}_j^0\hat{\mathbf{H}}_{jj}(\theta)+\mathbf{\Lambda}_m^0\hat{\mathbf{H}}_{mj}(\theta)\right]\left[\mathbf{D}_{\mathbf{\Lambda}_j^0\hat{\mathbf{H}}_{jj}+\mathbf{\Lambda}_m^0\hat{\mathbf{H}}_{mj}}(\theta)\right]^{-1}\left[\mathbf{\Lambda}_j^0\hat{\mathbf{H}}_{jj}(\theta)+\mathbf{\Lambda}_m^0\hat{\mathbf{H}}_{mj}(\theta)\right]' \\
&= N^{-1}\left[\hat{\mathbf{\Lambda}}_j(\theta)-\mathbf{\Lambda}_j^0\hat{\mathbf{H}}_{jj}(\theta)-\mathbf{\Lambda}_m^0\hat{\mathbf{H}}_{mj}(\theta)\right]\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_j}(\theta)\right]^{-1}\left[\hat{\mathbf{\Lambda}}_j(\theta)-\mathbf{\Lambda}_j^0\hat{\mathbf{H}}_{jj}(\theta)-\mathbf{\Lambda}_m^0\hat{\mathbf{H}}_{mj}(\theta)\right]' \\
&\quad +N^{-1}\left[\hat{\mathbf{\Lambda}}_j(\theta)-\mathbf{\Lambda}_j^0\hat{\mathbf{H}}_{jj}(\theta)-\mathbf{\Lambda}_m^0\hat{\mathbf{H}}_{mj}(\theta)\right]\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_j}(\theta)\right]^{-1}\left[\mathbf{\Lambda}_j^0\hat{\mathbf{H}}_{jj}(\theta)+\mathbf{\Lambda}_m^0\hat{\mathbf{H}}_{mj}(\theta)\right]' \\
&\quad +N^{-1}\left[\mathbf{\Lambda}_j^0\hat{\mathbf{H}}_{jj}(\theta)+\mathbf{\Lambda}_m^0\hat{\mathbf{H}}_{mj}(\theta)\right]\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_j}(\theta)\right]^{-1}\left[\hat{\mathbf{\Lambda}}_j(\theta)-\mathbf{\Lambda}_j^0\hat{\mathbf{H}}_{jj}(\theta)-\mathbf{\Lambda}_m^0\hat{\mathbf{H}}_{mj}(\theta)\right]' \\
&\quad +N^{-1}\left[\mathbf{\Lambda}_j^0\hat{\mathbf{H}}_{jj}(\theta)+\mathbf{\Lambda}_m^0\hat{\mathbf{H}}_{mj}(\theta)\right]\left\{\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_j}(\theta)\right]^{-1}-\left[\mathbf{D}_{\mathbf{\Lambda}_j^0\hat{\mathbf{H}}_{jj}+\mathbf{\Lambda}_m^0\hat{\mathbf{H}}_{mj}}(\theta)\right]^{-1}\right\} \\
&\quad \times\left[\mathbf{\Lambda}_j^0\hat{\mathbf{H}}_{jj}(\theta)+\mathbf{\Lambda}_m^0\hat{\mathbf{H}}_{mj}(\theta)\right]'
\end{aligned}.$$

We consider the case $j=1$: the case $j=2$ is analogous and omitted. We have

$$\begin{aligned}
& (NT)^{-1}\sum_{t=1}^{T}\mathbf{x}_t'\mathbb{I}_{1t}(\theta)\left[\mathbf{P}_{\hat{\mathbf{\Lambda}}_1}(\theta)-\mathbf{P}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right]\mathbf{x}_t \\
= & (NT)^{-1}\sum_{t=1}^{T}\mathbf{x}_t'\mathbb{I}_{1t}(\theta)N^{-1}\left[\hat{\mathbf{\Lambda}}_1(\theta)-\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}(\theta)-\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}(\theta)\right]\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}\left[\hat{\mathbf{\Lambda}}_1(\theta)-\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}(\theta)-\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}(\theta)\right]'\mathbf{x}_t \\
& +(NT)^{-1}\sum_{t=1}^{T}\mathbf{x}_t'\mathbb{I}_{1t}(\theta)N^{-1}\left[\hat{\mathbf{\Lambda}}_1(\theta)-\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}(\theta)-\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}(\theta)\right]\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}\left[\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}(\theta)+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}(\theta)\right]'\mathbf{x}_t \\
& +(NT)^{-1}\sum_{t=1}^{T}\mathbf{x}_t'\mathbb{I}_{1t}(\theta)N^{-1}\left[\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}(\theta)+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}(\theta)\right]\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}\left[\hat{\mathbf{\Lambda}}_1(\theta)-\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}(\theta)-\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}(\theta)\right]'\mathbf{x}_t \\
& +(NT)^{-1}\sum_{t=1}^{T}\mathbf{x}_t'\mathbb{I}_{1t}(\theta)N^{-1}\left[\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}(\theta)+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}(\theta)\right]\left\{\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}-\left[\mathbf{D}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right]^{-1}\right\}\left[\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}(\theta)+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}(\theta)\right]'\mathbf{x}_t \\
= & a_1(\theta)+a_2(\theta)+a_3(\theta)+a_4(\theta).
\end{aligned}$$

Starting from $a_1(\theta)$,

$$\begin{aligned}
a_1(\theta) &= N^{-2}\sum_{i=1}^{N}\sum_{l=1}^{N}\left\{\begin{array}{l}\left[\hat{\boldsymbol{\lambda}}_{1i}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right]'\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}\left[\hat{\boldsymbol{\lambda}}_{1l}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1l}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2l}^0\right] \\ \times\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)x_{it}x_{lt}\right]\end{array}\right\} \\
&\le \left\{N^{-2}\sum_{i=1}^{N}\sum_{l=1}^{N}\left\{\left[\hat{\boldsymbol{\lambda}}_{1i}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right]'\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}\left[\hat{\boldsymbol{\lambda}}_{1l}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1l}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2l}^0\right]\right\}^2\right\}^{1/2} \\
&\quad \times\left\{N^{-2}\sum_{i=1}^{N}\sum_{l=1}^{N}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)x_{it}x_{lt}\right]^2\right\}^{1/2} \\
&\le \left[N^{-1}\sum_{i=1}^{N}\left\|\hat{\boldsymbol{\lambda}}_{1i}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right\|^2\right]\left\|\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}\right\|O_p(1) \\
&= \left[N^{-1}\sum_{i=1}^{N}\left\|\hat{\boldsymbol{\lambda}}_{1i}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right\|^2\right]O_p(1)
\end{aligned}$$

by Lemma A.1(b) and the fact that $\left\|\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}\right\|=O_p(1)$, which is proved below: from Theorem 3.2 it follows that $a_1(\theta)=O_p\left(C_{NT}^{-2}\right)$ for all θ. As for $a_2(\theta)$,

$$\begin{aligned}
a_2(\theta) &= N^{-2}\sum_{i=1}^{N}\sum_{l=1}^{N}\left\{\left[\hat{\boldsymbol{\lambda}}_{1i}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right]'\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}\left[\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1l}^0+\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2l}^0\right]\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)x_{it}x_{lt}\right]\right\} \\
&\le \left\{N^{-2}\sum_{i=1}^{N}\sum_{l=1}^{N}\left\|\hat{\boldsymbol{\lambda}}_{1i}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right\|^2\left\|\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1l}^0+\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2l}^0\right\|^2\left\|\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}\right\|^2\right\}^{1/2} \\
&\quad \times\left\{N^{-2}\sum_{i=1}^{N}\sum_{l=1}^{N}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)x_{it}x_{lt}\right]^2\right\}^{1/2} \\
&\le \left[N^{-1}\sum_{i=1}^{N}\left\|\hat{\boldsymbol{\lambda}}_{1i}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right\|^2\right]^{1/2}\left\|\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}\right\|\left[N^{-1}\sum_{i=1}^{N}\left\|\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0+\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right\|^2\right]^{1/2}O_p(1) \\
&= \left[N^{-1}\sum_{i=1}^{N}\left\|\hat{\boldsymbol{\lambda}}_{1i}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right\|^2\right]^{1/2}O_p(1),
\end{aligned}$$

and $a_2(\theta) = O_p\left(C_{NT}^{-1}\right)$ for all θ. In an analogous way it is proved that $a_3(\theta) = O_p\left(C_{NT}^{-1}\right)$ for all θ. Finally,

$$
\begin{aligned}
a_4(\theta) &= N^{-2}\sum_{i=1}^{N}\sum_{l=1}^{N}\left\{\begin{array}{l}\left[\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0+\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right]'\left\{\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}-\left[\mathbf{D}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right]^{-1}\right\}\left[\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1l}^0+\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2l}^0\right]\\ \times\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)x_{it}x_{lt}\right]\end{array}\right\}\\
&\le \left\{N^{-2}\sum_{i=1}^{N}\sum_{l=1}^{N}\left\|\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0+\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right\|^2\left\|\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1l}^0+\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2l}^0\right\|^2\left\|\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}-\left[\mathbf{D}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right]^{-1}\right\|^2\right\}^{1/2}\\
&\quad\times\left\{N^{-2}\sum_{i=1}^{N}\sum_{l=1}^{N}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)x_{it}x_{lt}\right]^2\right\}^{1/2}\\
&\le \left\|\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}-\left[\mathbf{D}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right]^{-1}\right\|\left[N^{-1}\sum_{i=1}^{N}\left\|\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0+\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right\|^2+\right]O_p(1)\\
&\le \left\|\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}-\left[\mathbf{D}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right]^{-1}\right\|\left\{\left\|\hat{\mathbf{H}}_{11}(\theta)\right\|^2\left[N^{-1}\sum_{i=1}^{N}\left\|\boldsymbol{\lambda}_{1i}^0\right\|^2\right]+\left\|\hat{\mathbf{H}}_{21}(\theta)\right\|^2\left[N^{-1}\sum_{i=1}^{N}\left\|\boldsymbol{\lambda}_{2i}^0\right\|^2\right]\right\}O_p(1)\\
&= \left\|\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}-\left[\mathbf{D}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right]^{-1}\right\|O_p(1)
\end{aligned}
$$

where $O_p(1)$ comes from Lemma A.1(b) and Assumptions C1 and C2. Now, $\left\|\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)-\mathbf{D}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right\| = O_p\left(C_{NT}^{-1}\right)$ for all θ: this is because

$$
\begin{aligned}
\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)-\mathbf{D}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta) &= \frac{\hat{\mathbf{\Lambda}}_1(\theta)'\hat{\mathbf{\Lambda}}_1(\theta)}{N}-\frac{\left[\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}(\theta)+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}(\theta)\right]'\left[\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}(\theta)+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}(\theta)\right]}{N}\\
&= N^{-1}\sum_{i=1}^{N}\left\{\hat{\boldsymbol{\lambda}}_{1i}(\theta)\hat{\boldsymbol{\lambda}}_{1i}(\theta)'-\left[\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0+\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right]\left[\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0+\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right]'\right\}\\
&= N^{-1}\sum_{i=1}^{N}\left[\hat{\boldsymbol{\lambda}}_{1i}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right]\left[\hat{\boldsymbol{\lambda}}_{1i}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right]'\\
&\quad +N^{-1}\sum_{i=1}^{N}\left[\hat{\boldsymbol{\lambda}}_{1i}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right]\left[\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0+\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right]'\\
&\quad +N^{-1}\sum_{i=1}^{N}\left[\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0+\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right]\left[\hat{\boldsymbol{\lambda}}_{1i}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right]'
\end{aligned}
$$

so that

$$
\begin{aligned}
\left\|\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)-\mathbf{D}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right\| &\le N^{-1}\sum_{i=1}^{N}\left\|\hat{\boldsymbol{\lambda}}_{1i}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right\|^2\\
&\quad +2\left[N^{-1}\sum_{i=1}^{N}\left\|\hat{\boldsymbol{\lambda}}_{1i}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right\|^2\right]^{1/2}\\
&\quad \times\left[N^{-1}\sum_{i=1}^{N}\left\|\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0+\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right\|^2\right]^{1/2}\\
&= N^{-1}\sum_{i=1}^{N}\left\|\hat{\boldsymbol{\lambda}}_{1i}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right\|^2\\
&\quad +2\left[N^{-1}\sum_{i=1}^{N}\left\|\hat{\boldsymbol{\lambda}}_{1i}(\theta)-\hat{\mathbf{H}}_{11}(\theta)'\boldsymbol{\lambda}_{1i}^0-\hat{\mathbf{H}}_{21}(\theta)'\boldsymbol{\lambda}_{2i}^0\right\|^2\right]^{1/2}O_p(1)
\end{aligned}
$$

and the result follows. In general,

$$
\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}-\left[\mathbf{D}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right]^{-1}=\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)-\mathbf{D}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right]\left[\mathbf{D}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right]^{-1}
$$

and

$$
\left\|\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}-\left[\mathbf{D}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right]^{-1}\right\|\le\left\|\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)-\mathbf{D}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right\|\left\|\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}\right\|\left\|\left[\mathbf{D}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right]^{-1}\right\|.
$$

The matrix $\mathbf{\Lambda}_j^{0\prime}\mathbf{\Lambda}_j^0/N$ converges to a positive definite matrix by Assumption C2, for $j=1,2$, and the rank of $\hat{\mathbf{H}}_{11}(\theta)$ is equal to R^0 for all θ: since the rank of $\hat{\mathbf{H}}_{21}(\theta)$ is equal to R^0 for $\theta\neq\theta^0$, and $\hat{\mathbf{H}}_{21}(\theta^0)=\mathbf{0}_{R^0}$, this implies that $\mathbf{D}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)$ converges to a positive definite matrix. Since $\left\|\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)-\mathbf{D}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right\| = O_p\left(C_{NT}^{-1}\right)$, $\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)$

also converges to a positive definite matrix: this implies that $\left\|\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1}\right\| = O_p(1)$: therefore,

$$\begin{aligned}\left\|\left[\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta)\right]^{-1} - \left[\mathbf{D}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right]^{-1}\right\| &= \left\|\mathbf{D}_{\hat{\mathbf{\Lambda}}_1}(\theta) - \mathbf{D}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right\| O_p(1) \\ &= O_p\left(C_{NT}^{-1}\right)\end{aligned}$$

and $a_4(\theta) = O_p\left(C_{NT}^{-1}\right)$ for all θ. Combining all above results, we have

$$a_1(\theta) + a_2(\theta) + a_3(\theta) + a_4(\theta) = O_p\left(C_{NT}^{-2}\right) + O_p\left(C_{NT}^{-1}\right) + O_p\left(C_{NT}^{-1}\right) + O_p\left(C_{NT}^{-1}\right) = O_p\left(C_{NT}^{-1}\right):$$

this completes the proof of the lemma. ■

Proof of Lemma A.3. Let

$$\mathbf{P}_{\mathbf{\Lambda}_j^0} = \mathbf{\Lambda}_j^0\left(\mathbf{\Lambda}_j^{0\prime}\mathbf{\Lambda}_j^0\right)^{-1}\mathbf{\Lambda}_j^{0\prime}, \quad j = 1, 2,$$

and recall $\mathbf{P}_{\mathbf{\Lambda}_j^0\hat{\mathbf{H}}_{jj}+\mathbf{\Lambda}_m^0\hat{\mathbf{H}}_{mj}}(\theta)$ as defined in (22). Write

$$\begin{aligned}& S_{\mathbf{F}}\left[\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}(\theta) + \mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}(\theta), \mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}(\theta) + \mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{12}(\theta), \theta\right] - S_{\mathbf{F}}\left(\mathbf{\Lambda}_1^0, \mathbf{\Lambda}_2^0, \theta^0\right) \\ = & (NT)^{-1}\sum_{t=1}^{T}\mathbf{x}_t'\left\{\left[\mathbb{I}_{1t}\left(\theta^0\right)\mathbf{P}_{\mathbf{\Lambda}_1^0} - \mathbb{I}_{1t}(\theta)\mathbf{P}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right] + \left[\mathbb{I}_{2t}\left(\theta^0\right)\mathbf{P}_{\mathbf{\Lambda}_2^0} - \mathbb{I}_{2t}(\theta)\mathbf{P}_{\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}+\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{12}}(\theta)\right]\right\}\mathbf{x}_t \\ = & (NT)^{-1}\sum_{t=1}^{T}\left\{\begin{array}{l}\left[\mathbb{I}_{1t}\left(\theta^0\right)\mathbf{\Lambda}_1^0\mathbf{f}_t^0 + \mathbb{I}_{2t}\left(\theta^0\right)\mathbf{\Lambda}_2^0\mathbf{f}_t^0 + \mathbf{e}_t\right]' \\ \times\left\{\left[\mathbb{I}_{1t}\left(\theta^0\right)\mathbf{P}_{\mathbf{\Lambda}_1^0} - \mathbb{I}_{1t}(\theta)\mathbf{P}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right] + \left[\mathbb{I}_{2t}\left(\theta^0\right)\mathbf{P}_{\mathbf{\Lambda}_2^0} - \mathbb{I}_{2t}(\theta)\mathbf{P}_{\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}+\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{12}}(\theta)\right]\right\} \\ \times\left[\mathbb{I}_{1t}\left(\theta^0\right)\mathbf{\Lambda}_1^0\mathbf{f}_t^0 + \mathbb{I}_{2t}\left(\theta^0\right)\mathbf{\Lambda}_2^0\mathbf{f}_t^0 + \mathbf{e}_t\right]\end{array}\right\} \\ = & b_1(\theta) + b_2(\theta) + b_3(\theta),\end{aligned}$$

where

$$\begin{aligned}b_1(\theta) = & (NT)^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)\mathbb{I}_{1t}\left(\theta^0\right)\left[\mathbf{f}_t^{0\prime}\mathbf{\Lambda}_1^{0\prime}\mathbf{P}_{\mathbf{\Lambda}_1^0}\mathbf{\Lambda}_1^0\mathbf{f}_t^0 - \mathbf{f}_t^{0\prime}\mathbf{\Lambda}_1^{0\prime}\mathbf{P}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\mathbf{\Lambda}_1^0\mathbf{f}_t^0\right] \\ & + (NT)^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}(\theta)\mathbb{I}_{1t}\left(\theta^0\right)\left[\mathbf{f}_t^{0\prime}\mathbf{\Lambda}_1^{0\prime}\mathbf{P}_{\mathbf{\Lambda}_1^0}\mathbf{\Lambda}_1^0\mathbf{f}_t^0 - \mathbf{f}_t^{0\prime}\mathbf{\Lambda}_1^{0\prime}\mathbf{P}_{\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}+\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{12}}(\theta)\mathbf{\Lambda}_1^0\mathbf{f}_t^0\right] \\ & + (NT)^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)\mathbb{I}_{2t}\left(\theta^0\right)\left[\mathbf{f}_t^{0\prime}\mathbf{\Lambda}_2^{0\prime}\mathbf{P}_{\mathbf{\Lambda}_2^0}\mathbf{\Lambda}_2^0\mathbf{f}_t^0 - \mathbf{f}_t^{0\prime}\mathbf{\Lambda}_2^{0\prime}\mathbf{P}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\mathbf{\Lambda}_2^0\mathbf{f}_t^0\right] \\ & + (NT)^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}(\theta)\mathbb{I}_{2t}\left(\theta^0\right)\left[\mathbf{f}_t^{0\prime}\mathbf{\Lambda}_2^{0\prime}\mathbf{P}_{\mathbf{\Lambda}_2^0}\mathbf{\Lambda}_2^0\mathbf{f}_t^0 - \mathbf{f}_t^{0\prime}\mathbf{\Lambda}_2^{0\prime}\mathbf{P}_{\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}+\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{12}}(\theta)\mathbf{\Lambda}_2^0\mathbf{f}_t^0\right] \\ = & b_{11}(\theta) + b_{12}(\theta) + b_{13}(\theta) + b_{14}(\theta),\end{aligned}$$

$$\begin{aligned}b_2(\theta) = & 2(NT)^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)\mathbb{I}_{1t}\left(\theta^0\right)\left[\mathbf{e}_t'\mathbf{P}_{\mathbf{\Lambda}_1^0}\mathbf{\Lambda}_1^0\mathbf{f}_t^0 - \mathbf{e}_t'\mathbf{P}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\mathbf{\Lambda}_1^0\mathbf{f}_t^0\right] \\ & +2(NT)^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}(\theta)\mathbb{I}_{1t}\left(\theta^0\right)\left[\mathbf{e}_t'\mathbf{P}_{\mathbf{\Lambda}_1^0}\mathbf{\Lambda}_1^0\mathbf{f}_t^0 - \mathbf{e}_t'\mathbf{P}_{\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}+\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{12}}(\theta)\mathbf{\Lambda}_1^0\mathbf{f}_t^0\right] \\ & +2(NT)^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)\mathbb{I}_{2t}\left(\theta^0\right)\left[\mathbf{e}_t'\mathbf{P}_{\mathbf{\Lambda}_2^0}\mathbf{\Lambda}_2^0\mathbf{f}_t^0 - \mathbf{e}_t'\mathbf{P}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\mathbf{\Lambda}_2^0\mathbf{f}_t^0\right] \\ & +2(NT)^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}(\theta)\mathbb{I}_{2t}\left(\theta^0\right)\left[\mathbf{e}_t'\mathbf{P}_{\mathbf{\Lambda}_2^0}\mathbf{\Lambda}_2^0\mathbf{f}_t^0 - \mathbf{e}_t'\mathbf{P}_{\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}+\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{12}}(\theta)\mathbf{\Lambda}_2^0\mathbf{f}_t^0\right] \\ = & b_{21}(\theta) + b_{22}(\theta) + b_{23}(\theta) + b_{24}(\theta),\end{aligned}$$

and

$$
\begin{aligned}
b_3(\theta) &= (NT)^{-1}\sum_{t=1}^{T}\mathbf{e}_t'\left[\mathbb{I}_{1t}(\theta^0)-\mathbb{I}_{1t}(\theta)\right]\mathbf{P}_{\mathbf{\Lambda}_1^0}\mathbf{e}_t \\
&\quad+(NT)^{-1}\sum_{t=1}^{T}\mathbf{e}_t'\left[\mathbb{I}_{2t}(\theta^0)-\mathbb{I}_{2t}(\theta)\right]\mathbf{P}_{\mathbf{\Lambda}_2^0}\mathbf{e}_t \\
&\quad+(NT)^{-1}\sum_{t=1}^{T}\mathbf{e}_t'\mathbb{I}_{1t}(\theta)\left[\mathbf{P}_{\mathbf{\Lambda}_1^0}-\mathbf{P}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right]\mathbf{e}_t \\
&\quad+(NT)^{-1}\sum_{t=1}^{T}\mathbf{e}_t'\mathbb{I}_{2t}(\theta)\left[\mathbf{P}_{\mathbf{\Lambda}_2^0}-\mathbf{P}_{\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}+\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{12}}(\theta)\right]\mathbf{e}_t \\
&= b_{31}(\theta)+b_{32}(\theta)+b_{33}(\theta)+b_{34}(\theta).
\end{aligned}
$$

Consider $b_1(\theta)$ first. We have

$$
\begin{aligned}
b_{11}(\theta) &= \operatorname{tr}\left\{N^{-1}\left[\mathbf{\Lambda}_1^{0\prime}\mathbf{P}_{\mathbf{\Lambda}_1^0}\mathbf{\Lambda}_1^0-\mathbf{\Lambda}_1^{0\prime}\mathbf{P}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\mathbf{\Lambda}_1^0\right]\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)\mathbb{I}_{1t}(\theta^0)\mathbf{f}_t^0\mathbf{f}_t^{0\prime}\right]\right\} \\
&= \operatorname{tr}\left\{\left\{N^{-1}\mathbf{\Lambda}_1^{0\prime}\left[\mathbf{P}_{\mathbf{\Lambda}_1^0}-\mathbf{P}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right]\mathbf{\Lambda}_1^0\right\}\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)\mathbb{I}_{1t}(\theta^0)\mathbf{f}_t^0\mathbf{f}_t^{0\prime}\right]\right\} \\
&\xrightarrow{p} \operatorname{tr}\left\{\left\{\operatorname{p}\lim_{N\to\infty}\left\{N^{-1}\mathbf{\Lambda}_1^{0\prime}\left[\mathbf{P}_{\mathbf{\Lambda}_1^0}-\mathbf{P}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right]\mathbf{\Lambda}_1^0\right\}\right\}\mathbf{\Sigma}_{1\mathbf{f}}^0(\theta,\theta^0)\right\} \\
&= \operatorname{tr}\left[\mathbf{B}_{11}(\theta)\cdot\mathbf{\Sigma}_{1\mathbf{f}}^0(\theta,\theta^0)\right],
\end{aligned}
$$

where $\mathbf{B}_{11}(\theta)=\operatorname{p}\lim_{N\to\infty}\left\{N^{-1}\mathbf{\Lambda}_1^{0\prime}\left[\mathbf{P}_{\mathbf{\Lambda}_1^0}-\mathbf{P}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\right]\mathbf{\Lambda}_1^0\right\}$. Now $\mathbf{B}_{11}(\theta)$ is different from zero by Assumption C2 and it is also positive semi-definite. The matrix $\mathbf{\Sigma}_{1\mathbf{f}}^0(\theta,\theta^0)$ is positive definite by Assumption C1. It then follows that $\operatorname{p}\lim_{N,T\to\infty}b_{11}(\theta)=\operatorname{tr}\left[\mathbf{B}_{11}(\theta)\cdot\mathbf{\Sigma}_{1\mathbf{f}}^0(\theta,\theta^0)\right]>0$. Consider now

$$
\begin{aligned}
b_{12}(\theta) &= \operatorname{tr}\left\{N^{-1}\left[\mathbf{\Lambda}_1^{0\prime}\mathbf{P}_{\mathbf{\Lambda}_1^0}\mathbf{\Lambda}_1^0-\mathbf{\Lambda}_1^{0\prime}\mathbf{P}_{\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}+\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{12}}(\theta)\mathbf{\Lambda}_1^0\right]\left[T^{-1}\sum_{t=1}^{T}\mathbb{I}_{2t}(\theta)\mathbb{I}_{1t}(\theta^0)\mathbf{f}_t^0\mathbf{f}_t^{0\prime}\right]\right\} \\
&= \operatorname{tr}\left\{\left\{N^{-1}\mathbf{\Lambda}_1^{0\prime}\left[\mathbf{P}_{\mathbf{\Lambda}_1^0}-\mathbf{P}_{\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}+\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{12}}(\theta)\right]\mathbf{\Lambda}_1^0\right\}\left\{T^{-1}\sum_{t=1}^{T}\left[1-\mathbb{I}_{1t}(\theta)\right]\mathbb{I}_{1t}(\theta^0)\mathbf{f}_t^0\mathbf{f}_t^{0\prime}\right\}\right\} \\
&= \operatorname{tr}\left\{\left\{N^{-1}\mathbf{\Lambda}_1^{0\prime}\left[\mathbf{P}_{\mathbf{\Lambda}_1^0}-\mathbf{P}_{\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}+\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{12}}(\theta)\right]\mathbf{\Lambda}_1^0\right\}\left\{T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta^0)\mathbb{I}_{1t}(\theta^0)\mathbf{f}_t^0\mathbf{f}_t^{0\prime}-T^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)\mathbb{I}_{1t}(\theta^0)\mathbf{f}_t^0\mathbf{f}_t^{0\prime}\right\}\right\} \\
&\xrightarrow{p} \operatorname{tr}\left\{\operatorname{p}\lim_{N\to\infty}\left\{N^{-1}\mathbf{\Lambda}_1^{0\prime}\left[\mathbf{P}_{\mathbf{\Lambda}_1^0}-\mathbf{P}_{\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}+\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{12}}(\theta)\right]\mathbf{\Lambda}_1^0\right\}\left[\mathbf{\Sigma}_{1\mathbf{f}}^0(\theta^0,\theta^0)-\mathbf{\Sigma}_{1\mathbf{f}}^0(\theta,\theta^0)\right]\right\} \\
&= \operatorname{tr}\left\{\mathbf{B}_{12}(\theta)\left[\mathbf{\Sigma}_{1\mathbf{f}}^0(\theta^0,\theta^0)-\mathbf{\Sigma}_{1\mathbf{f}}^0(\theta,\theta^0)\right]\right\},
\end{aligned}
$$

where $\mathbf{B}_{12}(\theta)=\operatorname{p}\lim_{N\to\infty}\left\{N^{-1}\mathbf{\Lambda}_1^{0\prime}\left[\mathbf{P}_{\mathbf{\Lambda}_1^0}-\mathbf{P}_{\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}+\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{12}}(\theta)\right]\mathbf{\Lambda}_1^0\right\}$: taking into account Assumption C1, it follows that $\operatorname{p}\lim_{N,T\to\infty}b_{12}(\theta)\geq 0$. In a similar way it is proved that $\operatorname{p}\lim_{N,T\to\infty}b_{13}(\theta)\geq 0$ and $\operatorname{p}\lim_{N,T\to\infty}b_{14}(\theta)\geq 0$. Then

$$
\operatorname{p}\lim_{N\to\infty}b_1(\theta)=\operatorname{p}\lim_{N\to\infty}b_{11}(\theta)+\operatorname{p}\lim_{N\to\infty}b_{12}(\theta)+\operatorname{p}\lim_{N\to\infty}b_{13}(\theta)+\operatorname{p}\lim_{N\to\infty}b_{14}(\theta)>0,\quad\forall\theta\neq\theta^0.
$$

Consider now $b_2(\theta)$. We have

$$
b_{21}(\theta)=2(NT)^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)\mathbb{I}_{1t}(\theta^0)\mathbf{e}_t'\mathbf{P}_{\mathbf{\Lambda}_1^0}\mathbf{\Lambda}_1^0\mathbf{f}_t^0-2(NT)^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)\mathbb{I}_{1t}(\theta^0)\mathbf{e}_t'\mathbf{P}_{\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{21}}(\theta)\mathbf{\Lambda}_1^0\mathbf{f}_t^0.
$$

By Lemma A.1(c) and Assumption C1,

$$
\begin{aligned}
\left|(NT)^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)\mathbb{I}_{1t}(\theta^0)\mathbf{e}_t'\mathbf{P}_{\mathbf{\Lambda}_1^0}\mathbf{\Lambda}_1^0\mathbf{f}_t^0\right| &= \left|(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)\mathbb{I}_{1t}(\theta^0)e_{it}\boldsymbol{\lambda}_{1i}^{0\prime}\mathbf{f}_t^0\right| \\
&\leq \left[T^{-1}\sum_{t=1}^{T}\left\|\mathbb{I}_{1t}(\theta)\mathbf{f}_t^0\right\|^2\right]^{1/2}N^{-1/2}\left[N^{-1}\sum_{i=1}^{N}\left\|T^{-1/2}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta^0)e_{it}\boldsymbol{\lambda}_{1i}^0\right\|^2\right]^{1/2} \\
&= O_p\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}
$$

Further,

$$(NT)^{-1}\sum_{t=1}^{T}\mathbb{I}_{1t}(\theta)\,\mathbb{I}_{1t}\left(\theta^{0}\right)\mathbf{e}_{t}'\mathbf{P}_{\mathbf{\Lambda}_{1}^{0}\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_{2}^{0}\hat{\mathbf{H}}_{21}}(\theta)\,\mathbf{\Lambda}_{1}^{0}\mathbf{f}_{t}^{0}=O_{p}\left(\frac{1}{\sqrt{N}}\right).$$

Therefore, $b_{21}(\theta)=O_p\left(1\big/\sqrt{N}\right)$. In an analogous way, it can be proved that $b_{22}(\theta)=O_p\left(1\big/\sqrt{N}\right)$, $b_{23}(\theta)=O_p\left(1\big/\sqrt{N}\right)$, $b_{24}(\theta)=O_p\left(1\big/\sqrt{N}\right)$. Therefore, $b_{2}(\theta)=O_p\left(1\big/\sqrt{N}\right)\xrightarrow{p}0$ as $N\to\infty$.

Finally, consider $b_3(\theta)$. We have, $b_{31}(\theta)=o_p(1)$ and $b_{32}(\theta)=o_p(1)$. Further, $\left[\mathbf{P}_{\mathbf{\Lambda}_1^0}-\mathbf{P}_{\mathbf{\Lambda}_{1}^{0}\hat{\mathbf{H}}_{11}+\mathbf{\Lambda}_{2}^{0}\hat{\mathbf{H}}_{21}}(\theta)\right]$ and $\left[\mathbf{P}_{\mathbf{\Lambda}_2^0}-\mathbf{P}_{\mathbf{\Lambda}_{2}^{0}\hat{\mathbf{H}}_{22}+\mathbf{\Lambda}_{1}^{0}\hat{\mathbf{H}}_{12}}(\theta)\right]$ are positive semi-definite matrices, which implies that $b_{33}(\theta)\geq 0$ and $b_{34}(\theta)\geq 0$: this implies that $\operatorname{p\lim}_{N,T\to\infty}b_3(\theta)\geq 0$. This completes the proof of the lemma. ■

A.2 Proofs of Results in Section 3.5

Let

$$g_{it}^{0}(\theta_1,\theta_2)=\left|\mathbb{I}_{2t}(\theta_2)-\mathbb{I}_{2t}(\theta_1)\right|\left\|\mathbf{f}_t^0 e_{it}\right\|,\quad i=1,\ldots,N,\quad t=1,\ldots,T,$$

$$q_{t}^{0}(\theta_1,\theta_2)=\left|\mathbb{I}_{2t}(\theta_2)-\mathbb{I}_{2t}(\theta_1)\right|\left\|\mathbf{f}_t^0\right\|,\quad t=1,\ldots,T,$$

$$w_{it}^{0}(\theta)=\left|\mathbb{I}_{2t}(\theta)-\mathbb{I}_{2t}\left(\theta^0\right)\right|\left(\boldsymbol{\delta}_i^{0\prime}\mathbf{f}_t^0\right)^2,\quad i=1,\ldots,N,\quad t=1,\ldots,T,$$

$$w^{0}\left(\alpha^0,\theta\right)=\frac{1}{N^{\alpha^0}}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}w_{it}^{0}(\theta),$$

$$\mathbf{h}^{0}\left(\alpha^0,\theta\right)=\frac{1}{N^{\alpha^0}}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{I}_{2t}(\theta)\,\boldsymbol{\delta}_i^{0\prime}\mathbf{f}_t^0 e_{it}.$$

Lemma A.4 *There exists a $C_1<\infty$ such that for all $\theta_L\leq\theta_1\leq\theta_2\leq\theta_U$ and $s\leq 4$,*

$$\mathrm{E}\left\{\left[g_{it}^{0}(\theta_1,\theta_2)\right]^{s}\right\}\leq C_1\left|\theta_2-\theta_1\right|,\quad i=1,\ldots,N, \tag{23}$$

and

$$\mathrm{E}\left\{\left[q_{t}^{0}(\theta_1,\theta_2)\right]^{s}\right\}\leq C_1\left|\theta_2-\theta_1\right|. \tag{24}$$

Lemma A.5 *There exists a $K<\infty$ such that for all $\theta_L\leq\theta_1\leq\theta_2\leq\theta_U$,*

$$\mathrm{E}\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{\left[q_t^0(\theta_1,\theta_2)\right]^2-\mathrm{E}\left\{\left[q_t^0(\theta_1,\theta_2)\right]^2\right\}\right\}\right|^2\right]\leq K\left|\theta_2-\theta_1\right|.$$

Lemma A.6 *There exist constants $B>0$ and $0<d<\infty$ such that for all $\eta>0$ and $\varepsilon>0$, there exists a $\bar{v}<\infty$ such that for all N and T,*

$$\Pr\left[\inf_{\frac{\bar{v}}{N^{\alpha^0}T}\leq\left|\theta-\theta^0\right|\leq B}\frac{w^0\left(\alpha^0,\theta\right)}{\left|\theta-\theta^0\right|}<(1-\eta)\,d\right]\leq\varepsilon.$$

Lemma A.7 *For all $\eta>0$ and $\varepsilon>0$, there exists some $\bar{v}<\infty$ such that for any $B<\infty$,*

$$\Pr\left[\sup_{\frac{\bar{v}}{N^{\alpha^0}T}\leq\left|\theta-\theta^0\right|\leq B}\frac{\left\|\mathbf{h}^0\left(\alpha^0,\theta\right)-\mathbf{h}^0\left(\alpha^0,\theta^0\right)\right\|}{\left|\theta-\theta^0\right|}>\eta\right]\leq\varepsilon.$$

Proof of Theorem 3.4. Let B and d be defined as in Lemma A.6. Pick $\eta>0$ small enough so that

$$(1-\eta)\,d-2\eta>0. \tag{25}$$

Let $\mathbb{E}_{NT}$ be the joint event that $\left|\hat{\theta}-\theta^0\right|\leq B$, $\left\|\hat{\boldsymbol{\lambda}}_{ji}'\hat{\mathbf{f}}_t-\boldsymbol{\lambda}_{ji}^{0\prime}\mathbf{f}_t^0\right\|$ is small enough so that (28) below is satisfied, for $j=1,2$, $i=1,\ldots,N$, $t=1,\ldots,T$, and

$$\inf_{\frac{\bar{v}}{N^{\alpha^0}T}\leq|\theta-\theta^0|\leq B}\frac{w^0\left(\alpha^0,\theta\right)}{|\theta-\theta^0|}\geq(1-\eta)\,d, \tag{26}$$

and

$$\sup_{\frac{\bar{v}}{N^{\alpha^0}T}\leq|\theta-\theta^0|\leq B}\frac{\left\|\mathbf{h}^0\left(\alpha^0,\theta\right)-\mathbf{h}^0\left(\alpha^0,\theta^0\right)\right\|}{|\theta-\theta^0|}\leq\eta. \tag{27}$$

Fix $\varepsilon>0$ and pick $\bar{v}$, $\bar{N}$ and $\bar{T}$ so that $\Pr\left(\mathbb{E}_{NT}\right)\geq 1-\varepsilon$ for all $N\geq\bar{N}$ and $T\geq\bar{T}$, which is possible under Corollary 3.1, and Lemmas A.6 and A.7. Given $S\left(\boldsymbol{\Lambda},\mathbf{F},\theta\right)$ defined in (2), let

$$S\left(\alpha^0,\boldsymbol{\Lambda},\mathbf{F},\theta\right)=\frac{1}{N^{\alpha^0}T}\sum_{t=1}^{T}\left[\mathbf{x}_t-\boldsymbol{\Lambda}_1\mathbf{f}_t-\boldsymbol{\Delta}\mathbf{f}_{2t}\left(\theta\right)\right]'\left[\mathbf{x}_t-\boldsymbol{\Lambda}_1\mathbf{f}_t-\boldsymbol{\Delta}\mathbf{f}_{2t}\left(\theta\right)\right],$$

where $\mathbf{f}_{2t}\left(\theta\right)=\mathbb{I}_{2t}\left(\theta\right)\mathbf{f}_t$ and $\boldsymbol{\Delta}=\boldsymbol{\Lambda}_2-\boldsymbol{\Lambda}_1$. Since $S\left(\alpha^0,\boldsymbol{\Lambda},\mathbf{F},\theta\right)$ is continuous in $\left(\boldsymbol{\Lambda},\mathbf{F}\right)$, for small enough $\left\|\hat{\boldsymbol{\lambda}}_{ji}'\hat{\mathbf{f}}_t-\boldsymbol{\lambda}_{ji}^{0\prime}\mathbf{f}_t^0\right\|$, for $j=1,2$, $i=1,\ldots,N$, $t=1,\ldots,T$, it follows that

$$\begin{aligned}
S\left(\alpha^0,\hat{\boldsymbol{\Lambda}},\hat{\mathbf{F}},\theta\right)-S\left(\alpha^0,\hat{\boldsymbol{\Lambda}},\hat{\mathbf{F}},\theta^0\right) &= \frac{1}{N^{\alpha^0}T}\sum_{t=1}^{T}\left[\mathbf{x}_t-\hat{\boldsymbol{\Lambda}}_1\hat{\mathbf{f}}_t-\hat{\boldsymbol{\Delta}}\hat{\mathbf{f}}_{2t}\left(\theta\right)\right]'\left[\mathbf{x}_t-\hat{\boldsymbol{\Lambda}}_1\hat{\mathbf{f}}_t-\hat{\boldsymbol{\Delta}}\hat{\mathbf{f}}_{2t}\left(\theta\right)\right]\\
&\quad-\frac{1}{N^{\alpha^0}T}\sum_{t=1}^{T}\left[\mathbf{x}_t-\hat{\boldsymbol{\Lambda}}_1\hat{\mathbf{f}}_t-\hat{\boldsymbol{\Delta}}\hat{\mathbf{f}}_{2t}\left(\theta^0\right)\right]'\left[\mathbf{x}_t-\hat{\boldsymbol{\Lambda}}_1\hat{\mathbf{f}}_t-\hat{\boldsymbol{\Delta}}\hat{\mathbf{f}}_{2t}\left(\theta^0\right)\right]\\
&= D\left\{\begin{array}{l}\dfrac{1}{N^{\alpha^0}T}\sum_{t=1}^{T}\left[\mathbf{x}_t-\boldsymbol{\Lambda}_1^0\mathbf{f}_t^0-\boldsymbol{\Delta}^0\mathbf{f}_{2t}^0\left(\theta\right)\right]'\left[\mathbf{x}_t-\boldsymbol{\Lambda}_1^0\mathbf{f}_t^0-\boldsymbol{\Delta}^0\mathbf{f}_{2t}^0\left(\theta\right)\right]\\ -\dfrac{1}{N^{\alpha^0}T}\sum_{t=1}^{T}\left[\mathbf{x}_t-\boldsymbol{\Lambda}_1^0\mathbf{f}_t^0-\boldsymbol{\Delta}^0\mathbf{f}_{2t}^0\left(\theta^0\right)\right]'\left[\mathbf{x}_t-\boldsymbol{\Lambda}_1^0\mathbf{f}_t^0-\boldsymbol{\Delta}^0\mathbf{f}_{2t}^0\left(\theta^0\right)\right]\end{array}\right\}\\
&= D\left[S\left(\alpha^0,\boldsymbol{\Lambda}^0,\mathbf{F}^0,\theta\right)-S\left(\alpha^0,\boldsymbol{\Lambda}^0,\mathbf{F}^0,\theta^0\right)\right],
\end{aligned} \tag{28}$$

for some $D>0$, where $\hat{\mathbf{f}}_{2t}\left(\theta\right)=\mathbb{I}_{2t}\left(\theta\right)\hat{\mathbf{f}}_t$, $\hat{\boldsymbol{\Delta}}=\hat{\boldsymbol{\Lambda}}_2-\hat{\boldsymbol{\Lambda}}_1$, $\mathbf{f}_{2t}^0\left(\theta\right)=\mathbb{I}_{2t}\left(\theta\right)\mathbf{f}_t^0$ and $\boldsymbol{\Delta}^0=\boldsymbol{\Lambda}_2^0-\boldsymbol{\Lambda}_1^0$: the sign of $S\left(\alpha^0,\hat{\boldsymbol{\Lambda}},\hat{\mathbf{F}},\theta\right)-S\left(\alpha^0,\hat{\boldsymbol{\Lambda}},\hat{\mathbf{F}},\theta^0\right)$ is then equal to the sign of $S\left(\alpha^0,\boldsymbol{\Lambda}^0,\mathbf{F}^0,\theta\right)-S\left(\alpha^0,\boldsymbol{\Lambda}^0,\mathbf{F}^0,\theta^0\right)$. We have

$$\begin{aligned}
S\left(\alpha^0,\boldsymbol{\Lambda}^0,\mathbf{F}^0,\theta\right)-S\left(\alpha^0,\boldsymbol{\Lambda}^0,\mathbf{F}^0,\theta^0\right) &= \frac{1}{N^{\alpha^0}T}\sum_{t=1}^{T}\left[\mathbf{f}_{2t}^0\left(\theta\right)-\mathbf{f}_{2t}^0\left(\theta^0\right)\right]'\boldsymbol{\Delta}^{0\prime}\boldsymbol{\Delta}^0\left[\mathbf{f}_{2t}^0\left(\theta\right)-\mathbf{f}_{2t}^0\left(\theta^0\right)\right]\\
&\quad-2\frac{1}{N^{\alpha^0}T}\sum_{t=1}^{T}\left[\mathbf{f}_{2t}^0\left(\theta\right)-\mathbf{f}_{2t}^0\left(\theta^0\right)\right]'\boldsymbol{\Delta}^{0\prime}\mathbf{e}_t\\
&= S_1\left(\alpha^0,\theta\right)+S_2\left(\alpha^0,\theta\right)
\end{aligned}$$

and

$$\begin{aligned}
\frac{S\left(\alpha^0,\boldsymbol{\Lambda}^0,\mathbf{F}^0,\theta\right)-S\left(\alpha^0,\boldsymbol{\Lambda}^0,\mathbf{F}^0,\theta^0\right)}{|\theta-\theta^0|} &= \frac{1}{N^{\alpha^0}T\left|\theta-\theta^0\right|}\sum_{t=1}^{T}\left[\mathbf{f}_{2t}^0\left(\theta\right)-\mathbf{f}_{2t}^0\left(\theta^0\right)\right]'\boldsymbol{\Delta}^{0\prime}\boldsymbol{\Delta}^0\left[\mathbf{f}_{2t}^0\left(\theta\right)-\mathbf{f}_{2t}^0\left(\theta^0\right)\right]\\
&\quad-2\frac{1}{N^{\alpha^0}T\left|\theta-\theta^0\right|}\sum_{t=1}^{T}\left[\mathbf{f}_{2t}^0\left(\theta\right)-\mathbf{f}_{2t}^0\left(\theta^0\right)\right]'\boldsymbol{\Delta}^{0\prime}\mathbf{e}_t\\
&= \frac{S_1\left(\alpha^0,\theta\right)}{|\theta-\theta^0|}+\frac{S_2\left(\alpha^0,\theta\right)}{|\theta-\theta^0|}.
\end{aligned} \tag{29}$$

Suppose $\theta\in\left[\theta^0+\bar{v}N^{-\alpha^0}T^{-1},\theta^0+B\right]$ and that event $\mathbb{E}_{NT}$ holds. It follows that

$$\begin{aligned}
\frac{S_1\left(\alpha^0,\theta\right)}{\theta-\theta^0} &= \frac{1}{N^{\alpha^0}T\left(\theta-\theta^0\right)}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathbf{f}_{2t}^0\left(\theta\right)-\mathbf{f}_{2t}^0\left(\theta^0\right)\right]'\boldsymbol{\delta}_i^0\boldsymbol{\delta}_i^{0\prime}\left[\mathbf{f}_{2t}^0\left(\theta\right)-\mathbf{f}_{2t}^0\left(\theta^0\right)\right]\\
&= \frac{1}{N^{\alpha^0}T\left(\theta-\theta^0\right)}\sum_{i=1}^{N}\sum_{t=1}^{T}\left|\mathbb{I}_{2t}\left(\theta\right)-\mathbb{I}_{2t}\left(\theta^0\right)\right|\left(\boldsymbol{\delta}_i^{0\prime}\mathbf{f}_t^0\right)^2\\
&= \frac{w^0\left(\alpha^0,\theta\right)}{\theta-\theta^0},
\end{aligned} \tag{30}$$

and

$$
\begin{aligned}
\frac{S_2\left(\alpha^0,\theta\right)}{\theta-\theta^0} &= -2\frac{1}{N^{\alpha^0}T\left|\theta-\theta^0\right|}\sum_{t=1}^{T}\left[\mathbf{f}_{2t}^0\left(\theta\right)-\mathbf{f}_{2t}^0\left(\theta^0\right)\right]'\boldsymbol{\Delta}^{0\prime}\mathbf{e}_t \\
&= -2\frac{1}{N^{\alpha^0}T\left(\theta-\theta^0\right)}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathbf{f}_{2t}^0\left(\theta\right)-\mathbf{f}_{2t}^0\left(\theta^0\right)\right]'\boldsymbol{\delta}_i^0 e_{it} \\
&\geq -2\frac{1}{\theta-\theta^0}\left\|\frac{1}{N^{\alpha^0}T}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathbf{f}_{2t}^0\left(\theta\right)-\mathbf{f}_{2t}^0\left(\theta^0\right)\right]'\boldsymbol{\delta}_i^0 e_{it}\right\| \\
&= -2\frac{\left\|\mathbf{h}^0\left(\alpha^0,\theta\right)-\mathbf{h}^0\left(\alpha^0,\theta^0\right)\right\|}{\theta-\theta^0},
\end{aligned}
\tag{31}
$$

By (25) through (31) it follows that for some $D>0$,

$$
\frac{S\left(\alpha^0,\hat{\mathbf{\Lambda}},\hat{\mathbf{F}},\theta\right)-S\left(\alpha^0,\hat{\mathbf{\Lambda}},\hat{\mathbf{F}},\theta^0\right)}{\theta-\theta^0}\geq D\left[\frac{w^0\left(\alpha^0,\theta\right)}{\theta-\theta^0}-2\frac{\left\|\mathbf{h}^0\left(\alpha^0,\theta\right)-\mathbf{h}^0\left(\alpha^0,\theta^0\right)\right\|}{\theta-\theta^0}\right]\geq D\left[\left(1-\eta\right)d-2\eta\right]\geq 0.
$$

Given the event $\mathbb{E}_{NT}$, if $\theta\in\left[\theta^0+\bar{v}N^{-\alpha^0}T^{-1},\theta^0+B\right]$ then $S\left(\alpha^0,\hat{\mathbf{\Lambda}},\hat{\mathbf{F}},\theta\right)-S\left(\alpha^0,\hat{\mathbf{\Lambda}},\hat{\mathbf{F}},\theta^0\right)>0$. In a similar way, it can be shown that if $\theta\in\left[\theta^0-B,\theta^0-\bar{v}N^{-\alpha^0}T^{-1}\right]$ then $S\left(\alpha^0,\hat{\mathbf{\Lambda}},\hat{\mathbf{F}},\theta\right)-S\left(\alpha^0,\hat{\mathbf{\Lambda}},\hat{\mathbf{F}},\theta^0\right)>0$. As $S\left(\alpha^0,\hat{\mathbf{\Lambda}},\hat{\mathbf{F}},\hat{\theta}\right)-S\left(\alpha^0,\hat{\mathbf{\Lambda}},\hat{\mathbf{F}},\theta^0\right)\leq 0$, if $\mathbb{E}_{NT}$ occurs then $\left|\hat{\theta}-\theta^0\right|\leq\bar{v}N^{-\alpha^0}T^{-1}$: since $\Pr\left(\mathbb{E}_{NT}\right)\geq 1-\varepsilon$ for $N\geq\bar{N}$ and $T\geq\bar{T}$, then $\Pr\left(\left|\hat{\theta}-\theta^0\right|>\bar{v}N^{-\alpha^0}T^{-1}\right)\leq\varepsilon$ for $N\geq\bar{N}$ and $T\geq\bar{T}$: this is sufficient to show that $N^{\alpha^0}T\left(\hat{\theta}-\theta^0\right)=O_p\left(1\right)$. The convergence rate of the estimator for the loadings follows from (11). ■

Proof of Corollary 3.2. Corollary 3.2 easily follows from Theorem 3.4 and the proof is omitted. ■

Proof of Lemma A.4. We show (23): the proof of (24) is analogous. Given a random matrix $\mathbf{A}$,

$$
\frac{\partial}{\partial\theta}\mathrm{E}\left[\mathbf{A}\mathbb{I}_{1t}\left(\theta\right)\right]=\mathrm{E}\left(\mathbf{A}\left|z_t=\theta\right.\right)f_{\mathcal{Z}}\left(\theta\right). \tag{32}
$$

Under Assumption CR(b)

$$
\frac{\partial}{\partial\theta}\mathrm{E}\left[\left\|\mathbf{f}_t^0 e_{it}\right\|^s\mathbb{I}_{1t}\left(\theta\right)\right]=\mathrm{E}\left(\left\|\mathbf{f}_t^0 e_{it}\right\|^s\left|z_t=\theta\right.\right)f_{\mathcal{Z}}\left(\theta\right)\leq\left[\mathrm{E}\left(\left\|\mathbf{f}_t^0 e_{it}\right\|^4\left|z_t=\theta\right.\right)\right]^{s/4}f_{\mathcal{Z}}\left(\theta\right)\leq C^{s/4}\bar{f}\leq C_1,
$$

where $C_1=\max\left[1,C\right]\bar{f}$. For $\theta_1\leq\theta_2$, $\mathbb{I}_{1t}\left(\theta_2\right)-\mathbb{I}_{1t}\left(\theta_1\right)$ is either equal to one or to zero: by a first-order Taylor expansion, it follows that

$$
\mathrm{E}\left\{\left[g_{it}^0\left(\theta_1,\theta_2\right)\right]^s\right\}=\mathrm{E}\left[\left|\mathbb{I}_{2t}\left(\theta_2\right)-\mathbb{I}_{2t}\left(\theta_1\right)\right|\left\|\mathbf{f}_t^0 e_{it}\right\|^s\right]=\mathrm{E}\left\{\left[\mathbb{I}_{1t}\left(\theta_2\right)-\mathbb{I}_{1t}\left(\theta_1\right)\right]\left\|\mathbf{f}_t^0 e_{it}\right\|^s\right\}\leq C_1\left|\theta_2-\theta_1\right|.
$$

■

Proof of Lemma A.5. Lemma 3.4 in Peligrad (1982) shows that under Assumption CR(a) there exists a $K'<\infty$ such that, taking into account (24) in Lemma A.4,

$$
\begin{aligned}
\mathrm{E}\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{\left[q_t^0\left(\theta_1,\theta_2\right)\right]^2-\mathrm{E}\left\{\left[q_t^0\left(\theta_1,\theta_2\right)\right]^2\right\}\right\}\right|^2\right] &\leq K'\mathrm{E}\left\{\left\{\left[q_t^0\left(\theta_1,\theta_2\right)\right]^2-\mathrm{E}\left\{\left[q_t^0\left(\theta_1,\theta_2\right)\right]^2\right\}\right\}^2\right\} \\
&\leq 2K'\mathrm{E}\left\{\left[q_t^0\left(\theta_1,\theta_2\right)\right]^4\right\} \\
&\leq 2K'C_1\left|\theta_2-\theta_1\right|:
\end{aligned}
$$

setting $K=2K'C_1$ completes the proof of the lemma. ■

Proof of Lemma A.6. For $\theta \geq \theta^0$,

$$\begin{aligned}
\mathrm{E}\left[w^0\left(\alpha^0,\theta\right)\right] &= \frac{1}{N^{\alpha^0}}\sum_{i=1}^{N}\mathrm{E}\left[w_{it}^0\left(\theta\right)\right] \\
&= \frac{1}{N^{\alpha^0}}\sum_{i=1}^{N^{\alpha^0}}\mathrm{E}\left[w_{it}^0\left(\theta\right)\right] + \frac{1}{N^{\alpha^0}}\sum_{i=N^{\alpha^0}+1}^{N}\mathrm{E}\left[w_{it}^0\left(\theta\right)\right] \\
&= \frac{1}{N^{\alpha^0}}\sum_{i=1}^{N^{\alpha^0}}\boldsymbol{\delta}_i^{0\prime}\left[\boldsymbol{\Sigma}_{2\mathbf{f}}^0\left(\theta^0,\theta^0\right)-\boldsymbol{\Sigma}_{2\mathbf{f}}^0\left(\theta,\theta\right)\right]\boldsymbol{\delta}_i^0 + \frac{1}{N^{\alpha^0}}\sum_{i=N^{\alpha^0}+1}^{N}\boldsymbol{\delta}_i^{0\prime}\left[\boldsymbol{\Sigma}_{2\mathbf{f}}^0\left(\theta^0,\theta^0\right)-\boldsymbol{\Sigma}_{2\mathbf{f}}^0\left(\theta,\theta\right)\right]\boldsymbol{\delta}_i^0,
\end{aligned}$$

and

$$\frac{\partial\mathrm{E}\left[w^0\left(\alpha^0,\theta\right)\right]}{\partial\theta} = \frac{1}{N^{\alpha^0}}\sum_{i=1}^{N^{\alpha^0}}\boldsymbol{\delta}_i^{0\prime}\mathbf{D}_{\mathbf{f}}^0\left(\theta\right)f_{\mathcal{Z}}\left(\theta\right)\boldsymbol{\delta}_i^0 + \frac{1}{N^{\alpha^0}}\sum_{i=N^{\alpha^0}+1}^{N}\boldsymbol{\delta}_i^{0\prime}\mathbf{D}_{\mathbf{f}}^0\left(\theta\right)f_{\mathcal{Z}}\left(\theta\right)\boldsymbol{\delta}_i^0$$

by (32) (the sign is reversed if $\theta \leq \theta^0$). By Assumptions CR(c) and CR(d), $\partial\mathrm{E}\left[w^0\left(\alpha^0,\theta\right)\right]/\partial\theta$ is continuous at $\theta=\theta^0$, and $\partial\mathrm{E}\left[w^0\left(\alpha^0,\theta^0\right)\right]/\partial\theta>0$, respectively: there then exists a B small enough such that for $\left|\theta-\theta^0\right|\leq B$

$$d = \min_{\left|\theta-\theta^0\right|\leq B}\frac{\partial\mathrm{E}\left[w^0\left(\alpha^0,\theta\right)\right]}{\partial\theta} > 0.$$

The first-order Taylor expansion of $\mathrm{E}\left[w^0\left(\alpha^0,\theta\right)\right]$ about $\theta=\theta^0$ results in

$$\inf_{\left|\theta-\theta^0\right|\leq B}\mathrm{E}\left[w^0\left(\alpha^0,\theta\right)\right] \geq d\left|\theta-\theta^0\right|, \tag{33}$$

since $\mathrm{E}\left[w^0\left(\alpha^0,\theta^0\right)\right]=0$. Without loss of generality, set $\boldsymbol{\delta}_i^0=\mathbf{0}$, for $i=N^{\alpha^0}+1,\ldots,N$. Notice that

$$\begin{aligned}
\mathrm{E}\left\{\left|w^0\left(\alpha^0,\theta\right)-\mathrm{E}\left[w^0\left(\alpha^0,\theta\right)\right]\right|^2\right\} &= \mathrm{E}\left\{\left|\frac{1}{N^{\alpha^0}}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{w_{it}^0\left(\theta\right)-\mathrm{E}\left[w_{it}^0\left(\theta\right)\right]\right\}\right|^2\right\} \\
&= \mathrm{E}\left\{\left|\frac{1}{N^{\alpha^0}}\frac{1}{T}\sum_{i=1}^{N^{\alpha^0}}\sum_{t=1}^{T}\left\{w_{it}^0\left(\theta\right)-\mathrm{E}\left[w_{it}^0\left(\theta\right)\right]\right\}\right|^2\right\} \\
&\leq \frac{C_2}{N^{2\alpha^0}}\sum_{i=1}^{N^{\alpha^0}}\mathrm{E}\left\{\left|\frac{1}{T}\sum_{t=1}^{T}\left\{w_{it}^0\left(\theta\right)-\mathrm{E}\left[w_{it}^0\left(\theta\right)\right]\right\}\right|^2\right\}
\end{aligned}$$

for some $C_2<\infty$, and

$$\begin{aligned}
\mathrm{E}\left\{\left|\frac{1}{T}\sum_{t=1}^{T}\left\{w_{it}^0\left(\theta\right)-\mathrm{E}\left[w_{it}^0\left(\theta\right)\right]\right\}\right|^2\right\} &\leq \left\|\boldsymbol{\delta}_i^0\right\|^4T^{-1}\mathrm{E}\left\{\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{q_t^0\left(\theta,\theta^0\right)-\mathrm{E}\left[q_t^0\left(\theta,\theta^0\right)\right]\right\}\right|^2\right\}, \quad i=1,\ldots,N^{\alpha^0}, \\
&\leq \left\|\boldsymbol{\delta}_i^0\right\|^4T^{-1}K\left|\theta-\theta^0\right|
\end{aligned}$$

by Lemma A.5: since

$$\left\|\boldsymbol{\delta}_i^0\right\| = \left\|\boldsymbol{\lambda}_{2i}^0-\boldsymbol{\lambda}_{1i}^0\right\| \leq \left\|\boldsymbol{\lambda}_{1i}^0\right\|+\left\|\boldsymbol{\lambda}_{2i}^0\right\| \leq 2\bar{\lambda}, \quad i=1,\ldots,N^{\alpha^0}, \tag{34}$$

by Assumption C2, it follows that

$$\mathrm{E}\left\{\left|w^0\left(\alpha^0,\theta\right)-\mathrm{E}\left[w^0\left(\alpha^0,\theta\right)\right]\right|^2\right\} \leq \frac{C_2 16\bar{\lambda}^4}{N^{\alpha^0}T}K\left|\theta-\theta^0\right|. \tag{35}$$

For any η and ε, set

$$b = \frac{1-\eta/2}{1-\eta} > 1 \tag{36}$$

and

$$\bar{v} = \frac{8C_2 16\bar{\lambda}^4K}{\eta^2d^2\left(1-1/b\right)^2\varepsilon}. \tag{37}$$

Assume N and T large enough so that $\bar{v}\Big/\left(N^{\alpha^0}T\right) \leq B$, otherwise the lemma is trivially satisfied. For $l_N = 1,\ldots,N+1$ and $l_T = 1,\ldots,T+1$, set $\theta_{l_N l_T} = \theta^0 + \bar{v}b^{l_N-1}b^{l_T-1}\Big/\left(N^{\alpha^0}T\right)$, where N and T are integers such that $\theta_{NT}-\theta^0 = \bar{v}b^{N-1}b^{T-1}\Big/\left(N^{\alpha^0}T\right) \leq B$, $\theta_{N+1,T}-\theta^0 > B$ and $\theta_{N,T+1}-\theta^0 > B$ (since $\bar{v}\Big/\left(N^{\alpha^0}T\right) \leq B$ then $NT \geq 1$). By Markov's inequality, (33), (35) and (37),

$$\begin{aligned}
\Pr\left\{\sup_{\substack{1\leq l_N\leq N,\\ 1\leq l_T\leq T}}\left|\frac{w^0\left(\alpha^0,\theta_{l_N l_T}\right)}{\mathrm{E}\left[w^0\left(\alpha^0,\theta_{l_N l_T}\right)\right]}-1\right|>\frac{\eta}{2}\right\} &\leq \left(\frac{2}{\eta}\right)^2\sum_{l_N=1}^{N}\sum_{l_T=1}^{T}\frac{\mathrm{E}\left\{\left|w^0\left(\alpha^0,\theta_{l_N l_T}\right)-\mathrm{E}\left[w^0\left(\alpha^0,\theta_{l_N l_T}\right)\right]\right|^2\right\}}{\left|\mathrm{E}\left[w^0\left(\alpha^0,\theta_{l_N l_T}\right)\right]\right|^2}\\
&\leq \frac{4}{\eta^2}\sum_{l_N=1}^{N}\sum_{l_T=1}^{T}\frac{C_2 N^{-\alpha^0}T^{-1}16\bar{\lambda}^4 K\left(\theta_{l_N l_T}-\theta^0\right)}{d^2\left(\theta_{l_N l_T}-\theta^0\right)^2}\\
&\leq \frac{4}{\eta^2}\frac{C_2 16\bar{\lambda}^4 K}{d^2\bar{v}}\left(\sum_{l_N=0}^{\infty}\frac{1}{b^{l_N}}\right)\left(\sum_{l_T=0}^{\infty}\frac{1}{b^{l_T}}\right)\\
&= \frac{4}{\eta^2}\frac{C_2 16\bar{\lambda}^4 K}{d^2\bar{v}}\frac{1}{\left(1-1/b\right)^2}\leq\frac{\varepsilon}{2}:
\end{aligned}$$

it follows that for all $1 \leq l_N \leq N$ and $1 \leq l_T \leq T$, and with probability greater than $1-\varepsilon/2$,

$$\left|\frac{w^0\left(\alpha^0,\theta_{l_N l_T}\right)}{\mathrm{E}\left[w^0\left(\alpha^0,\theta_{l_N l_T}\right)\right]}-1\right| \leq \frac{\eta}{2}. \tag{38}$$

Using (36), for any θ such that $\bar{v}\Big/\left(N^{\alpha^0}T\right) \leq \left(\theta-\theta^0\right) \leq B$, there exists some $l_N \leq N$ and $l_T \leq T$ such that $\theta_{l_N l_T} < \theta < \min\left\{\theta_{l_{N+1},l_T},\theta_{l_N,l_{T+1}}\right\}$ and on the event (38)

$$\frac{w^0\left(\alpha^0,\theta\right)}{\left(\theta-\theta^0\right)} \geq \frac{w^0\left(\alpha^0,\theta_{l_N l_T}\right)}{\mathrm{E}\left[w^0\left(\alpha^0,\theta_{l_N l_T}\right)\right]}\frac{\mathrm{E}\left[w^0\left(\alpha^0,\theta_{l_N l_T}\right)\right]}{\left[\min\left\{\theta_{l_{N+1},l_T},\theta_{l_N,l_{T+1}}\right\}-\theta^0\right]} \geq \left(1-\frac{\eta}{2}\right)\frac{d\left(\theta_{l_N l_T}-\theta^0\right)}{\left[\min\left\{\theta_{l_{N+1},l_T},\theta_{l_N,l_{T+1}}\right\}-\theta^0\right]} = (1-\eta)\,d$$

where we set $\left(\theta_{l_N l_T}-\theta^0\right)\Big/\left[\min\left\{\theta_{l_{N+1},l_T},\theta_{l_N,l_{T+1}}\right\}-\theta^0\right] = 1/b$: this event has probability greater than $1-\varepsilon/2$ and then

$$\Pr\left[\inf_{\frac{\bar{v}}{N^{\alpha^0}T}\leq\left(\theta-\theta^0\right)\leq B}\frac{w^0\left(\alpha^0,\theta\right)}{\left(\theta-\theta^0\right)} < (1-\eta)\,d\right] \leq \frac{\varepsilon}{2},$$

holds. Taking the infimum over $-\bar{v}\Big/\left(N^{\alpha^0}T\right) \geq \left(\theta-\theta^0\right) \geq -B$ allows to prove a similar inequality using the same argument: this completes the proof of the lemma. ■

Proof of Lemma A.7. Given some $C_3 < \infty$ to be determined later, fix $\eta > 0$ and set

$$\bar{v} = \frac{8}{(0.5)^2\,(0.5)^2}\frac{C_1 C_3\bar{\lambda}^2}{\eta^2\varepsilon}. \tag{39}$$

For $l_N = 1, \ldots, N$ and $l_T = 1, \ldots, T$, set $\theta_{l_N l_T} - \theta^0 = \bar{v} 2^{l_N - 1} 2^{l_T - 1} \Big/ \left(N^{\alpha^0} T \right) \leq B$. Without loss of generality, assume that $\boldsymbol{\delta}_i^0 = \mathbf{0}$, for $i = N^{\alpha^0} + 1, \ldots, N$. Markov's inequality, (23) in Lemma A.4, (34) and (39) ensure that

$$\begin{aligned}
\Pr\left[\sup_{\substack{1 \leq l_N \leq N, \\ 1 \leq l_T \leq T}} \frac{\left\| \mathbf{h}^0\left(\alpha^0, \theta_{l_N l_T}\right) - \mathbf{h}^0\left(\alpha^0, \theta^0\right) \right\|}{\left(\theta_{l_N l_T} - \theta^0\right)} > \eta \right] &\leq \frac{1}{\eta^2} \sum_{l_N=1}^{N} \sum_{l_T=1}^{T} \frac{\mathrm{E}\left[\left\| \mathbf{h}^0\left(\alpha^0, \theta_{l_N l_T}\right) - \mathbf{h}^0\left(\alpha^0, \theta^0\right) \right\|^2 \right]}{\left(\theta_{l_N l_T} - \theta^0\right)^2} \\
&\leq \frac{1}{\eta^2} \sum_{l_N=1}^{N} \sum_{l_T=1}^{T} \frac{\mathrm{E}\left\{ \left\| \frac{1}{N^{\alpha^0}} \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\mathbb{I}_{2t}\left(\theta_{l_N l_T}\right) - \mathbb{I}_{2t}\left(\theta^0\right) \right] \boldsymbol{\delta}_i^{0\prime} \mathbf{f}_t^0 e_{it} \right\|^2 \right\}}{\left(\theta_{l_N l_T} - \theta^0\right)^2} \\
&\leq \frac{C_3}{\eta^2} \frac{1}{N^{2\alpha^0}} \frac{1}{T^2} \sum_{l_N=1}^{N} \sum_{l_T=1}^{T} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{\mathrm{E}\left\{ \left\| \left[\mathbb{I}_{2t}\left(\theta_{l_N l_T}\right) - \mathbb{I}_{2t}\left(\theta^0\right) \right] \boldsymbol{\delta}_i^{0\prime} \mathbf{f}_t^0 e_{it} \right\|^2 \right\}}{\left(\theta_{l_N l_T} - \theta^0\right)^2} \\
&\leq \frac{C_3}{\eta^2} \frac{1}{N^{2\alpha^0}} \frac{1}{T} \sum_{l_N=1}^{N} \sum_{l_T=1}^{T} \sum_{i=1}^{N} \frac{\left\| \boldsymbol{\delta}_i^0 \right\|^2 \mathrm{E}\left\{ \left[\mathbb{I}_{2t}\left(\theta_{l_N l_T}\right) - \mathbb{I}_{2t}\left(\theta^0\right) \right] \left\| \mathbf{f}_t^0 e_{it} \right\|^2 \right\}}{\left(\theta_{l_N l_T} - \theta^0\right)^2} \\
&= \frac{C_3}{\eta^2} \frac{1}{N^{2\alpha^0}} \frac{1}{T} \sum_{l_N=1}^{N} \sum_{l_T=1}^{T} \sum_{i=1}^{N^{\alpha^0}} \frac{\left\| \boldsymbol{\delta}_i^0 \right\|^2 \mathrm{E}\left\{ \left[\mathbb{I}_{2t}\left(\theta_{l_N l_T}\right) - \mathbb{I}_{2t}\left(\theta^0\right) \right] \left\| \mathbf{f}_t^0 e_{it} \right\|^2 \right\}}{\left(\theta_{l_N l_T} - \theta^0\right)^2} \\
&\leq \frac{C_3}{\eta^2} \frac{1}{N^{\alpha^0}} \frac{1}{T} \sum_{l_N=1}^{N} \sum_{l_T=1}^{T} \left[4\bar{\lambda}^2 \frac{C_1\left(\theta_{l_N l_T} - \theta^0\right)}{\left(\theta_{l_N l_T} - \theta^0\right)^2} \right] \\
&= 4 \frac{C_1 C_3 \bar{\lambda}^2}{\eta^2 \bar{v}} \left[\sum_{l_N=1}^{N} \frac{1}{\left(2^{l_N - 1}\right)} \right] \left[\sum_{l_T=1}^{T} \frac{1}{\left(2^{l_T - 1}\right)} \right] \\
&\leq \frac{4}{(0.5)^2 (0.5)^2} \frac{C_1 C_3 \bar{\lambda}^2}{\eta^2 \bar{v}} \leq \frac{\varepsilon}{2}.
\end{aligned}$$

It follows that for all $1 \leq l_N \leq N$ and $1 \leq l_T \leq T$, and with probability greater than $1 - \varepsilon/2$,

$$\frac{\left\| \mathbf{h}^0\left(\alpha^0, \theta_{l_N l_T}\right) - \mathbf{h}^0\left(\alpha^0, \theta^0\right) \right\|}{\left(\theta_{l_N l_T} - \theta^0\right)} \leq \eta,$$

which implies that

$$\Pr\left[\sup_{\frac{\bar{v}}{N^{\alpha^0} T} \leq \left(\theta - \theta^0\right) \leq B} \frac{\left\| \mathbf{h}^0\left(\alpha^0, \theta\right) - \mathbf{h}^0\left(\alpha^0, \theta^0\right) \right\|}{\left(\theta - \theta^0\right)} > \eta \right] \leq \frac{\varepsilon}{2}.$$

Taking the infimum over $-\bar{v} \Big/ \left(N^{\alpha^0} T \right) \geq \left(\theta - \theta^0\right) \geq -B$ allows to prove a similar inequality using the same argument, which completes the proof. ■

A.3 Proof of the Result in Section 4

Given the loss function in (12) and for any fixed $R \geq 1$, let $\hat{\boldsymbol{\Lambda}}_j^R(\theta) = \left[\hat{\boldsymbol{\lambda}}_{j1}^R(\theta), \ldots, \hat{\boldsymbol{\lambda}}_{jN}^R(\theta) \right]'$ be the $N \times R$ matrix of estimated loadings for fixed θ, for $j = 1, 2$. Let $\hat{\mathbf{V}}_j^R(\theta)$ be the $R \times R$ diagonal matrix of the first R largest eigenvalues of $\hat{\boldsymbol{\Sigma}}_{j\mathbf{x}}(\theta)$ in (7) in decreasing order, for $j = 1, 2$. Define the $R^0 \times R$ rotation matrix

$$\hat{\mathbf{H}}_{jj}^R\left(\theta^0\right) = \frac{\mathbf{F}_j^0\left(\theta^0\right) \mathbf{F}_j^0\left(\theta^0\right)'}{T} \frac{\boldsymbol{\Lambda}_j^{0\prime} \hat{\boldsymbol{\Lambda}}_j^R\left(\theta^0\right)}{N} \hat{\mathbf{V}}_j^R\left(\theta^0\right)^{-1}, \quad j = 1, 2, \tag{40}$$

where $\mathbf{F}_j^0(\theta)$ is defined in Section 3.4.

Lemma A.8 *For any fixed $R \geq 1$, there exists a $R^0 \times R$ matrix $\hat{\mathbf{H}}_{jj}^R\left(\theta^0\right)$ as defined in* (40)*, with* $\operatorname{rank}\left[\hat{\mathbf{H}}_{jj}^R\left(\theta^0\right) \right] = \min\left\{ R^0, R \right\}$*, and $C_{NT} = \min\left\{ \sqrt{N}, \sqrt{T} \right\}$, such that*

$$C_{NT}^2 \left[\frac{1}{N} \sum_{i=1}^{N} \left\| \hat{\boldsymbol{\lambda}}_{ji}^R\left(\theta^0\right) - \hat{\mathbf{H}}_{jj}^R\left(\theta^0\right)' \boldsymbol{\lambda}_{ji}^0 \right\|^2 \right] = O_p(1), \quad j = 1, 2.$$

Lemma A.9 *Let $\hat{\theta}^{\bar{R}}$ be the estimator for θ^0 obtained from the loss function in* (12) *for any a priori chosen number of factors $R = \bar{R}$ such that $\bar{R} \geq R^0$. Then under assumptions I, C1-C4 and CR,*

$$N^{\alpha^0} T \left(\hat{\theta}^{\bar{R}} - \theta^0 \right) = O_p (1) .$$

Proof of Theorem 4.1. Consider

$$\mathbf{x}_t = \mathbb{I}_{1t} \left(\theta^0\right) \mathbf{\Lambda}_1^0 \mathbf{f}_t^0 + \mathbb{I}_{2t} \left(\theta^0\right) \mathbf{\Lambda}_2^0 \mathbf{f}_t^0 + \mathbf{e}_t = \mathbf{\Lambda}^0 \left[\mathbb{I}_{1t} \left(\theta^0\right) \mathbf{f}_t^{0\prime}, \mathbb{I}_{2t} \left(\theta^0\right) \mathbf{f}_t^{0\prime}\right]' + \mathbf{e}_t,$$

where $\mathbf{\Lambda}^0 = \left(\mathbf{\Lambda}_1^0, \mathbf{\Lambda}_2^0\right) = \left[\left(\boldsymbol{\lambda}_{11}^0, \ldots, \boldsymbol{\lambda}_{1N}^0\right)', \left(\boldsymbol{\lambda}_{21}^0, \ldots, \boldsymbol{\lambda}_{2N}^0\right)'\right] = \left(\boldsymbol{\lambda}_1^0, \ldots, \boldsymbol{\lambda}_N^0\right)'$ is a $N \times 2R^0$ matrix, with $\boldsymbol{\lambda}_i^0 = \left(\boldsymbol{\lambda}_{1i}^{0\prime}, \boldsymbol{\lambda}_{2i}^{0\prime}\right)'$ a $2R^0 \times 1$ vector, and $\left[\mathbb{I}_{1t} \left(\theta^0\right) \mathbf{f}_t^{0\prime}, \mathbb{I}_{2t} \left(\theta^0\right) \mathbf{f}_t^{0\prime}\right]'$ is a $2R^0 \times 1$ vector. Given the loss function in (12), let $\hat{\mathbf{f}}_t^R (\theta)$ be the $R \times 1$ vector of estimated factors for fixed θ, for $t = 1, \ldots, T$. Further, let $\hat{\mathbf{H}}_{jj}^{R+} \left(\theta^0\right)$ be the generalized inverse of $\hat{\mathbf{H}}_{jj}^{R} \left(\theta^0\right)$ in (40) such that $\hat{\mathbf{H}}_{jj}^{R} \left(\theta^0\right) \hat{\mathbf{H}}_{jj}^{R+} \left(\theta^0\right) = \mathbf{I}_R$, for $j = 1, 2$. Lemma A.8 implies that

$$C_{NT}^2 \left\{ \frac{1}{T} \sum_{t=1}^{T} \left\| \hat{\mathbf{f}}_t^R \left(\theta^0\right) - \left[\mathbb{I}_{1t} \left(\theta^0\right) \hat{\mathbf{H}}_{11}^{R+} \left(\theta^0\right) + \mathbb{I}_{2t} \left(\theta^0\right) \hat{\mathbf{H}}_{22}^{R+} \left(\theta^0\right) \right] \mathbf{f}_t^0 \right\|^2 \right\} = O_p (1)$$

or

$$C_{NT}^2 \left\{ \frac{1}{T} \sum_{t=1}^{T} \left\| \begin{bmatrix} \mathbb{I}_{1t} \left(\theta^0\right) \hat{\mathbf{f}}_t^R \left(\theta^0\right) \\ \mathbb{I}_{2t} \left(\theta^0\right) \hat{\mathbf{f}}_t^R \left(\theta^0\right) \end{bmatrix} - \begin{bmatrix} \mathbb{I}_{1t} \left(\theta^0\right) \hat{\mathbf{H}}_{11}^{R+} \left(\theta^0\right) \mathbf{f}_t^0 \\ \mathbb{I}_{2t} \left(\theta^0\right) \hat{\mathbf{H}}_{22}^{R+} \left(\theta^0\right) \mathbf{f}_t^0 \end{bmatrix} \right\|^2 \right\} = O_p (1) ,$$

so that by Lemma A.9

$$C_{NT}^2 \left\{ \frac{1}{T} \sum_{t=1}^{T} \left\| \begin{bmatrix} \mathbb{I}_{1t} \left(\hat{\theta}^{\bar{R}}\right) \hat{\mathbf{f}}_t^R \left(\hat{\theta}^{\bar{R}}\right) \\ \mathbb{I}_{2t} \left(\hat{\theta}^{\bar{R}}\right) \hat{\mathbf{f}}_t^R \left(\hat{\theta}^{\bar{R}}\right) \end{bmatrix} - \begin{bmatrix} \mathbb{I}_{1t} \left(\theta^0\right) \hat{\mathbf{H}}_{11}^{R+} \left(\theta^0\right) \mathbf{f}_t^0 \\ \mathbb{I}_{2t} \left(\theta^0\right) \hat{\mathbf{H}}_{22}^{R+} \left(\theta^0\right) \mathbf{f}_t^0 \end{bmatrix} \right\|^2 \right\} = O_p (1) ,$$

which is analogous to Theorem 1 and Corollary 2 in Bai and Ng (2002): this is sufficient to complete the proof of the theorem, as it shows that the criteria in (13) select $\left(R^0 + R^0\right)$ factors. ■

Proof of Lemma A.8. The proof of Lemma A.8 is similar to that of Theorem 3.2 and omitted. ■

Proof of Lemma A.9. Given the loss function in (12) and following similar steps as in the proof of Theorem 3.3, it can be shown that

$$\lim_{N,T \to \infty} \mathrm{P} \left\{ S \left[\hat{\mathbf{\Lambda}}^R (\theta), \hat{\mathbf{F}}^R (\theta), \theta \right] \leq S \left[\hat{\mathbf{\Lambda}}^R \left(\theta^0\right), \hat{\mathbf{F}}^R \left(\theta^0\right), \theta^0 \right] \right\} = 0, \quad \forall \theta \neq \theta^0, \quad R^0 \leq R \leq R^{\max}.$$

In order to prove the lemma it is then sufficient to show that

$$S \left[\hat{\mathbf{\Lambda}}^R \left(\theta^0\right), \hat{\mathbf{F}}^R \left(\theta^0\right), \theta^0 \right] - S \left[\hat{\mathbf{\Lambda}} \left(\theta^0\right), \hat{\mathbf{F}} \left(\theta^0\right), \theta^0 \right] = O_p \left(C_{NT}^{-2} \right)$$

for any fixed R such that $R^0 \leq R \leq R^{\max}$, where $S \left[\hat{\mathbf{\Lambda}} (\theta), \hat{\mathbf{F}} (\theta), \theta \right] = S \left[\hat{\mathbf{\Lambda}}^{R^0} (\theta), \hat{\mathbf{F}}^{R^0} (\theta), \theta \right]$. Notice that

$$\begin{aligned} & \left| S \left[\hat{\mathbf{\Lambda}}^R \left(\theta^0\right), \hat{\mathbf{F}}^R \left(\theta^0\right), \theta^0 \right] - S \left[\hat{\mathbf{\Lambda}} \left(\theta^0\right), \hat{\mathbf{F}} \left(\theta^0\right), \theta^0 \right] \right| \\ \leq & \left| S \left[\hat{\mathbf{\Lambda}}^R \left(\theta^0\right), \hat{\mathbf{F}}^R \left(\theta^0\right), \theta^0 \right] - S \left(\mathbf{\Lambda}^0, \mathbf{F}^0, \theta^0 \right) \right| + \left| S \left(\mathbf{\Lambda}^0, \mathbf{F}^0, \theta^0 \right) - S \left[\hat{\mathbf{\Lambda}} \left(\theta^0\right), \hat{\mathbf{F}} \left(\theta^0\right), \theta^0 \right] \right| \\ \leq & \ 2 \max_{R^0 \leq R \leq R^{\max}} \left| S \left[\hat{\mathbf{\Lambda}}^R \left(\theta^0\right), \hat{\mathbf{F}}^R \left(\theta^0\right), \theta^0 \right] - S \left(\mathbf{\Lambda}^0, \mathbf{F}^0, \theta^0 \right) \right| : \end{aligned}$$

it therefore is sufficient to show that

$$S\left[\hat{\mathbf{\Lambda}}^R\left(\theta^0\right),\hat{\mathbf{F}}^R\left(\theta^0\right),\theta^0\right]-S\left(\mathbf{\Lambda}^0,\mathbf{F}^0,\theta^0\right)=O_p\left(C_{NT}^{-2}\right)$$

for each R such that $R^0\leq R\leq R^{\max}$. We have

$$\begin{aligned}\mathbf{x}_t &= \mathbb{I}_{1t}\left(\theta^0\right)\mathbf{\Lambda}_1^0\mathbf{f}_t^0+\mathbb{I}_{2t}\left(\theta^0\right)\mathbf{\Lambda}_2^0\mathbf{f}_t^0+\mathbf{e}_t\\ &= \mathbb{I}_{1t}\left(\theta^0\right)\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}^R\left(\theta^0\right)\hat{\mathbf{H}}_{11}^{R+}\left(\theta^0\right)\mathbf{f}_t^0+\mathbb{I}_{2t}\left(\theta^0\right)\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}^R\left(\theta^0\right)\hat{\mathbf{H}}_{22}^{R+}\left(\theta^0\right)\mathbf{f}_t^0+\mathbf{e}_t,\end{aligned}$$

where $\hat{\mathbf{H}}_{jj}^{R+}\left(\theta^0\right)$ is defined in the proof of Theorem 4.1, for $j=1,2$. This implies

$$\begin{aligned}\mathbf{x}_t &= \mathbb{I}_{1t}\left(\theta^0\right)\hat{\mathbf{\Lambda}}_1^R\left(\theta^0\right)\hat{\mathbf{H}}_{11}^{R+}\left(\theta^0\right)\mathbf{f}_t^0+\mathbb{I}_{2t}\left(\theta^0\right)\hat{\mathbf{\Lambda}}_2^R\left(\theta^0\right)\hat{\mathbf{H}}_{22}^{R+}\left(\theta^0\right)\mathbf{f}_t^0\\ &\quad+\mathbf{e}_t-\mathbb{I}_{1t}\left(\theta^0\right)\left[\hat{\mathbf{\Lambda}}_1^R\left(\theta^0\right)-\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}^R\left(\theta^0\right)\right]\hat{\mathbf{H}}_{11}^{R+}\left(\theta^0\right)\mathbf{f}_t^0-\mathbb{I}_{2t}\left(\theta^0\right)\left[\hat{\mathbf{\Lambda}}_2^R\left(\theta^0\right)-\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}^R\left(\theta^0\right)\right]\hat{\mathbf{H}}_{22}^{R+}\left(\theta^0\right)\mathbf{f}_t^0\\ &= \mathbb{I}_{1t}\left(\theta^0\right)\hat{\mathbf{\Lambda}}_1^R\left(\theta^0\right)\hat{\mathbf{H}}_{11}^{R+}\left(\theta^0\right)\mathbf{f}_t^0+\mathbb{I}_{2t}\left(\theta^0\right)\hat{\mathbf{\Lambda}}_2^R\left(\theta^0\right)\hat{\mathbf{H}}_{22}^{R+}\left(\theta^0\right)\mathbf{f}_t^0+\mathbf{u}_t,\end{aligned}$$

where

$$\mathbf{u}_t=\mathbf{e}_t-\mathbb{I}_{1t}\left(\theta^0\right)\left[\hat{\mathbf{\Lambda}}_1^R\left(\theta^0\right)-\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}^R\left(\theta^0\right)\right]\hat{\mathbf{H}}_{11}^{R+}\left(\theta^0\right)\mathbf{f}_t^0-\mathbb{I}_{2t}\left(\theta^0\right)\left[\hat{\mathbf{\Lambda}}_2^R\left(\theta^0\right)-\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}^R\left(\theta^0\right)\right]\hat{\mathbf{H}}_{22}^{R+}\left(\theta^0\right)\mathbf{f}_t^0.$$

Notice that

$$S\left(\mathbf{\Lambda}^0,\mathbf{F}^0,\theta^0\right)=(NT)^{-1}\sum_{t=1}^T\mathbf{e}_t'\mathbf{e}_t$$

and

$$\begin{aligned}&S\left[\hat{\mathbf{\Lambda}}^R\left(\theta^0\right),\hat{\mathbf{F}}^R\left(\theta^0\right),\theta^0\right]\\ =\;&(NT)^{-1}\sum_{t=1}^T\mathbf{u}_t'\mathbf{u}_t\\ =\;&(NT)^{-1}\sum_{t=1}^T\mathbf{e}_t'\mathbf{e}_t\\ &-2\,(NT)^{-1}\sum_{t=1}^T\mathbf{f}_t^{0\prime}\left\{\begin{array}{c}\mathbb{I}_{1t}\left(\theta^0\right)\hat{\mathbf{H}}_{11}^{R+}\left(\theta^0\right)'\left[\hat{\mathbf{\Lambda}}_1^R\left(\theta^0\right)-\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}^R\left(\theta^0\right)\right]'\\ +\mathbb{I}_{2t}\left(\theta^0\right)\hat{\mathbf{H}}_{22}^{R+}\left(\theta^0\right)'\left[\hat{\mathbf{\Lambda}}_2^R\left(\theta^0\right)-\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}^R\left(\theta^0\right)\right]'\end{array}\right\}\mathbf{e}_t\\ &+(NT)^{-1}\sum_{t=1}^T\mathbf{f}_t^{0\prime}\left\{\begin{array}{c}\mathbb{I}_{1t}\left(\theta^0\right)\hat{\mathbf{H}}_{11}^{R+}\left(\theta^0\right)'\left[\hat{\mathbf{\Lambda}}_1^R\left(\theta^0\right)-\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}^R\left(\theta^0\right)\right]'\left[\hat{\mathbf{\Lambda}}_1^R\left(\theta^0\right)-\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}^R\left(\theta^0\right)\right]\hat{\mathbf{H}}_{11}^{R+}\left(\theta^0\right)\\ +\mathbb{I}_{2t}\left(\theta^0\right)\hat{\mathbf{H}}_{22}^{R+}\left(\theta^0\right)'\left[\hat{\mathbf{\Lambda}}_2^R\left(\theta^0\right)-\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}^R\left(\theta^0\right)\right]'\left[\hat{\mathbf{\Lambda}}_2^R\left(\theta^0\right)-\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}^R\left(\theta^0\right)\right]\hat{\mathbf{H}}_{22}^{R+}\left(\theta^0\right)\end{array}\right\}\mathbf{f}_t^0\\ =\;&S\left(\mathbf{\Lambda}^0,\mathbf{F}^0,\theta^0\right)+S^{(1)}\left[\hat{\mathbf{\Lambda}}^R\left(\theta^0\right),\hat{\mathbf{F}}^R\left(\theta^0\right),\theta^0\right]+S^{(2)}\left[\hat{\mathbf{\Lambda}}^R\left(\theta^0\right),\hat{\mathbf{F}}^R\left(\theta^0\right),\theta^0\right]\end{aligned}$$

so that

$$\begin{aligned}\left|S\left[\hat{\mathbf{\Lambda}}^R\left(\theta^0\right),\hat{\mathbf{F}}^R\left(\theta^0\right),\theta^0\right]-S\left(\mathbf{\Lambda}^0,\mathbf{F}^0,\theta^0\right)\right| &= \left|S^{(1)}\left[\hat{\mathbf{\Lambda}}^R\left(\theta^0\right),\hat{\mathbf{F}}^R\left(\theta^0\right),\theta^0\right]+S^{(2)}\left[\hat{\mathbf{\Lambda}}^R\left(\theta^0\right),\hat{\mathbf{F}}^R\left(\theta^0\right),\theta^0\right]\right|\\ &\leq \left|S^{(1)}\left[\hat{\mathbf{\Lambda}}^R\left(\theta^0\right),\hat{\mathbf{F}}^R\left(\theta^0\right),\theta^0\right]\right|+\left|S^{(2)}\left[\hat{\mathbf{\Lambda}}^R\left(\theta^0\right),\hat{\mathbf{F}}^R\left(\theta^0\right),\theta^0\right]\right|.\end{aligned}$$

For any $A \times A$ matrix $\mathbf{A}$, $|\mathrm{tr}(\mathbf{A})| \leq A\|\mathbf{A}\|$. It follows that

$$
\begin{aligned}
& \left|S^{(1)}\left[\hat{\mathbf{\Lambda}}^R(\theta^0), \hat{\mathbf{F}}^R(\theta^0), \theta^0\right]\right| \\
= & \left|\mathrm{tr}\left\{S^{(1)}\left[\hat{\mathbf{\Lambda}}^R(\theta^0), \hat{\mathbf{F}}^R(\theta^0), \theta^0\right]\right\}\right| \\
= & 2\left|\mathrm{tr}\left\{(NT)^{-1}\sum_{t=1}^T \mathbf{f}_t^{0\prime}\left\{\begin{array}{c} \mathbb{I}_{1t}(\theta^0)\hat{\mathbf{H}}_{11}^{R+}(\theta^0)'\left[\hat{\mathbf{\Lambda}}_1^R(\theta^0)-\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}^R(\theta^0)\right]' \\ +\mathbb{I}_{2t}(\theta^0)\hat{\mathbf{H}}_{22}^{R+}(\theta^0)'\left[\hat{\mathbf{\Lambda}}_2^R(\theta^0)-\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}^R(\theta^0)\right]' \end{array}\right\}\mathbf{e}_t\right\}\right| \\
\leq & 2R\left[\begin{array}{c} \left\|\hat{\mathbf{H}}_{11}^{R+}(\theta^0)\right\|\left\|\dfrac{\hat{\mathbf{\Lambda}}_1^R(\theta^0)-\mathbf{\Lambda}_1^0\hat{\mathbf{H}}_{11}^R(\theta^0)}{\sqrt{N}}\right\|\left\|\dfrac{1}{\sqrt{N}T}\sum_{t=1}^T\mathbb{I}_{1t}(\theta^0)\mathbf{e}_t\mathbf{f}_t^{0\prime}\right\| \\ +\left\|\hat{\mathbf{H}}_{22}^{R+}(\theta^0)\right\|\left\|\dfrac{\hat{\mathbf{\Lambda}}_2^R(\theta^0)-\mathbf{\Lambda}_2^0\hat{\mathbf{H}}_{22}^R(\theta^0)}{\sqrt{N}}\right\|\left\|\dfrac{1}{\sqrt{N}T}\sum_{t=1}^T\mathbb{I}_{2t}(\theta^0)\mathbf{e}_t\mathbf{f}_t^{0\prime}\right\| \end{array}\right] \\
\leq & 2R\left\{\begin{array}{c} \left\|\hat{\mathbf{H}}_{11}^{R+}(\theta^0)\right\|\left[\dfrac{1}{N}\sum_{i=1}^N\left\|\hat{\boldsymbol{\lambda}}_{1i}^R(\theta^0)-\hat{\mathbf{H}}_{11}^R(\theta^0)'\boldsymbol{\lambda}_{1i}^0\right\|^2\right]^{1/2}\dfrac{1}{\sqrt{T}}\left[\dfrac{1}{N}\sum_{i=1}^N\left\|\dfrac{1}{\sqrt{T}}\sum_{t=1}^T\mathbb{I}_{1t}(\theta^0)e_{it}\mathbf{f}_t^0\right\|^2\right]^{1/2} \\ +\left\|\hat{\mathbf{H}}_{22}^{R+}(\theta^0)\right\|\left[\dfrac{1}{N}\sum_{i=1}^N\left\|\hat{\boldsymbol{\lambda}}_{2i}^R(\theta^0)-\hat{\mathbf{H}}_{22}^R(\theta^0)'\boldsymbol{\lambda}_{2i}^0\right\|^2\right]^{1/2}\dfrac{1}{\sqrt{T}}\left[\dfrac{1}{N}\sum_{i=1}^N\left\|\dfrac{1}{\sqrt{T}}\sum_{t=1}^T\mathbb{I}_{2t}(\theta^0)e_{it}\mathbf{f}_t^0\right\|^2\right]^{1/2} \end{array}\right\} \\
= & O_p\left(C_{NT}^{-1}\right)\frac{1}{\sqrt{T}}+O_p\left(C_{NT}^{-1}\right)\frac{1}{\sqrt{T}}=O_p\left(C_{NT}^{-2}\right)
\end{aligned}
$$

by Assumption C.4 and Lemma A.8. Further, by Lemma A.8

$$
\begin{aligned}
\left|S^{(2)}\left[\hat{\mathbf{\Lambda}}^R(\theta^0), \hat{\mathbf{F}}^R(\theta^0), \theta^0\right]\right| = & S^{(2)}\left[\hat{\mathbf{\Lambda}}^R(\theta^0), \hat{\mathbf{F}}^R(\theta^0), \theta^0\right] \\
\leq & \left[\frac{1}{N}\sum_{i=1}^N\left\|\hat{\boldsymbol{\lambda}}_{1i}^R(\theta^0)-\hat{\mathbf{H}}_{11}^R(\theta^0)'\boldsymbol{\lambda}_{1i}^0\right\|^2\right]\left\|\hat{\mathbf{H}}_{11}^{R+}(\theta^0)\right\|^2\left[\frac{1}{T}\sum_{t=1}^T\mathbb{I}_{1t}(\theta^0)\|\mathbf{f}_t^0\|^2\right] \\
& +\left[\frac{1}{N}\sum_{i=1}^N\left\|\hat{\boldsymbol{\lambda}}_{2i}^R(\theta^0)-\hat{\mathbf{H}}_{22}^R(\theta^0)'\boldsymbol{\lambda}_{2i}^0\right\|^2\right]\left\|\hat{\mathbf{H}}_{22}^{R+}(\theta^0)\right\|^2\left[\frac{1}{T}\sum_{t=1}^T\mathbb{I}_{2t}(\theta^0)\|\mathbf{f}_t^0\|^2\right] \\
\leq & \left(\frac{1}{T}\sum_{t=1}^T\|\mathbf{f}_t^0\|^2\right)\left\{\begin{array}{c} \left[\dfrac{1}{N}\sum_{i=1}^N\left\|\hat{\boldsymbol{\lambda}}_{1i}^R(\theta^0)-\hat{\mathbf{H}}_{11}^R(\theta^0)'\boldsymbol{\lambda}_{1i}^0\right\|^2\right]\left\|\hat{\mathbf{H}}_{11}^{R+}(\theta^0)\right\|^2 \\ +\left[\dfrac{1}{N}\sum_{i=1}^N\left\|\hat{\boldsymbol{\lambda}}_{2i}^R(\theta^0)-\hat{\mathbf{H}}_{22}^R(\theta^0)'\boldsymbol{\lambda}_{2i}^0\right\|^2\right]\left\|\hat{\mathbf{H}}_{22}^{R+}(\theta^0)\right\|^2 \end{array}\right\} \\
= & O_p(1)\cdot\left[O_p\left(C_{NT}^{-2}\right)\cdot O_p(1)+O_p\left(C_{NT}^{-2}\right)\cdot O_p(1)\right]=O_p\left(C_{NT}^{-2}\right),
\end{aligned}
$$

which completes the proof of the lemma. ■

A.4 Proofs of the Result in Section 5.2

Under Assumption LT1, $\left(T^{-1}\sum_{t=1}^T\left\|\tilde{\mathbf{f}}_t-\tilde{\mathbf{H}}_1^{-1}\mathbf{f}_t^0\right\|^2\right)=O_p\left(C_{NT}^{-2}\right)$, with $\tilde{\mathbf{H}}_1$ as in Theorem 3.1. Let $\tilde{\mathbf{h}}_1^{+\prime}$ be the first $1\times R^0$ row vector of $\tilde{\mathbf{H}}_1^{-1}$; and $\tilde{\mathbf{H}}_1^+$ be the $(R^0-1)\times R^0$ matrix containing the second to last row of $\tilde{\mathbf{H}}_1^{-1}$. We have

$$
\tilde{\mathbf{f}}_t=\begin{pmatrix}\tilde{f}_{1t} \\ \tilde{\mathbf{f}}_{-1,t}\end{pmatrix}\xrightarrow{p}\tilde{\mathbf{H}}_1^{-1}\mathbf{f}_t^0=\begin{pmatrix}\tilde{\mathbf{h}}_1^{+\prime}\mathbf{f}_t^0 \\ \tilde{\mathbf{H}}_1^+\mathbf{f}_t^0\end{pmatrix}=\begin{pmatrix}\tilde{f}_{1t}^{+0} \\ \tilde{\mathbf{f}}_{-1,t}^{+0}\end{pmatrix}=\tilde{\mathbf{f}}_t^{+0}.
$$

Define $\tilde{\mathbf{f}}_{-,t}^{+0}(\theta)=\left[\mathbb{I}_{1t}(\theta)\tilde{\mathbf{f}}_{-1,t}^{+0\prime},\mathbb{I}_{2t}(\theta)\tilde{\mathbf{f}}_{-1,t}^{+0\prime}\right]'$.

Lemma A.10 *For each θ,*

$$
\left\|\frac{1}{T}\sum_{t=1}^T\mathbb{I}_{jt}(\theta)\tilde{\mathbf{f}}_t\tilde{\mathbf{f}}_t'-\frac{1}{T}\sum_{t=1}^T\mathbb{I}_{jt}(\theta)\tilde{\mathbf{f}}_t^{+0}\tilde{\mathbf{f}}_t^{+0\prime}\right\|=O_p\left(C_{NT}^{-2}\right),\quad j=1,2.
$$

Lemma A.11 *For each θ,*

$$
\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T\tilde{\mathbf{f}}_{-,t}(\theta)\tilde{f}_{1t}-\frac{1}{\sqrt{T}}\sum_{t=1}^T\tilde{\mathbf{f}}_{-,t}^{+0}(\theta)\tilde{f}_{1t}^{+0}\right\|=o_p(1).
$$

Proof of Theorem 5.1. From Theorem 3.1, $\tilde{\mathbf{H}}_1 = \left(\mathbf{F}^0\mathbf{F}^{0\prime}/T\right)\left(\mathbf{\Lambda}_1^{0\prime}\tilde{\mathbf{\Lambda}}_1/N\right)\tilde{\mathbf{V}}_1^{-1}$. By Assumption LT2, $\left(\mathbf{F}^0\mathbf{F}^{0\prime}/T\right)\xrightarrow{p}\mathbf{\Sigma}_{\mathbf{f}}^0$. Following arguments similar to Proposition 1 in Bai (2003), $\left(\mathbf{\Lambda}_1^{0\prime}\tilde{\mathbf{\Lambda}}_1/N\right)\xrightarrow{p}\mathbf{Q}_{\mathbf{\Lambda}_1}$, where $\mathbf{Q}_{\mathbf{\Lambda}_1}$ is an invertible matrix and it is unique by Assumption LT3. By Lemma A.3 in Bai (2003), $\tilde{\mathbf{V}}_1\xrightarrow{p}\mathbf{V}_1$ where $\mathbf{V}_1$ is a positive definite matrix. It follows that $\tilde{\mathbf{H}}_1\xrightarrow{p}\mathbf{H}_1^0 = \mathbf{\Sigma}_{\mathbf{f}}^0\mathbf{Q}_{\mathbf{\Lambda}_1}\mathbf{V}_1^{-1}$, where $\mathbf{H}_1^0$ is an $R^0\times R^0$ invertible matrix and it is unique by Assumption LT3. Let $\mathbf{h}_1^{+0\prime}$ be the first $1\times R^0$ row vector of $\left(\mathbf{H}_1^0\right)^{-1}$. Let $\mathbf{H}_1^{+0}$ be the $\left(R^0-1\right)\times R^0$ matrix containing the second to last row of $\left(\mathbf{H}_1^0\right)^{-1}$. Define $f_{1t}^{+0} = \mathbf{h}_1^{+0\prime}\mathbf{f}_t^0$, $\mathbf{f}_{-1,t}^{+0} = \mathbf{H}_1^{+0}\mathbf{f}_t^0$ and $\mathbf{f}_{-,t}^{+0}(\theta) = \left[\mathbb{I}_{1t}(\theta)\mathbf{f}_{-1,t}^{+0\prime}, \mathbb{I}_{2t}(\theta)\mathbf{f}_{-1,t}^{+0\prime}\right]'$. From Lemma A.11 it follows that

$$\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T\tilde{\mathbf{f}}_{-,t}(\theta)\tilde{f}_{1t} - \frac{1}{\sqrt{T}}\sum_{t=1}^T\mathbf{f}_{-,t}^{+0}(\theta)f_{1t}^{+0}\right\| = o_p(1).$$

In order to prove that $\hat{\mathbf{k}}_-(\theta)\Rightarrow\mathbf{k}_-^0(\theta)$ it is sufficient to prove that $T^{-1/2}\sum_{t=1}^T\mathbf{f}_{-,t}^{+0}(\theta)f_{1t}^{+0}\Rightarrow\mathbf{k}_-^0(\theta)$: this follows if $T^{-1/2}\sum_{t=1}^T\mathbf{f}_{-,t}^{+0}(\theta)f_{1t}^{+0}$ is stochastically equicontinuous. As in Hansen (1996), we resort to Application 4 of Theorem 1 in Doukhan *et al.* (1995). Under Assumption LT5(a), the summands $\mathbf{k}_{-,t}^{+0}(\theta) = \mathbf{f}_{-,t}^{+0}(\theta)f_{1t}^{+0}$ satisfy the required $\beta-$mixing decay rate. Since $\left\|\mathbf{H}_1^{+0}\right\| = O(1)$ and $\left\|\mathbf{h}_1^{+0}\right\| = O(1)$, the envelope function $\sup_\theta\left\|\mathbf{k}_{-,t}^{+0}(\theta)\right\|$ satisfies

$$\begin{aligned}
\sup_\theta\left\|\mathbf{k}_{-,t}^{+0}(\theta)\right\| &= \sup_\theta\left\|\mathbf{f}_{-,t}^{+0}(\theta)f_{1t}^{+0}\right\| \\
&= \sup_\theta\left\|\left[\mathbb{I}_{1t}(\theta)\mathbf{f}_{-1,t}^{+0\prime}, \mathbb{I}_{2t}(\theta)\mathbf{f}_{-1,t}^{+0\prime}\right]'\mathbf{h}_1^{+0\prime}\mathbf{f}_t^0\right\| \\
&= \sup_\theta\left\|\left\{\left[\mathbb{I}_{1t}(\theta)\mathbf{H}_1^{+0}\mathbf{f}_t^0\right]', \left[\mathbb{I}_{2t}(\theta)\mathbf{H}_1^{+0}\mathbf{f}_t^0\right]'\right\}'\mathbf{h}_1^{+0\prime}\left[\mathbb{I}_{1t}(\theta)\mathbf{f}_t^0 + \mathbb{I}_{2t}(\theta)\mathbf{f}_t^0\right]\right\| \\
&\le \sup_\theta\left\|\left[\mathbb{I}_{1t}(\theta)\mathbf{f}_t^{0\prime}, \mathbb{I}_{2t}(\theta)\mathbf{f}_t^{0\prime}\right]'\left[\mathbb{I}_{1t}(\theta)\mathbf{f}_t^0 + \mathbb{I}_{2t}(\theta)\mathbf{f}_t^0\right]'\right\|O(1) \\
&\le \sup_\theta\left[\left\|\mathbb{I}_{1t}(\theta)\mathbf{f}_t^0\mathbf{f}_t^{0\prime}\right\| + \left\|\mathbb{I}_{2t}(\theta)\mathbf{f}_t^0\mathbf{f}_t^{0\prime}\right\|\right]O(1) \\
&\le \left[\sup_\theta\left\|\mathbb{I}_{1t}(\theta)\mathbf{f}_t^0\right\|\sup_\theta\left\|\mathbb{I}_{1t}(\theta)\mathbf{f}_t^0\right\| + \sup_\theta\left\|\mathbb{I}_{2t}(\theta)\mathbf{f}_t^0\right\|\sup_\theta\left\|\mathbb{I}_{2t}(\theta)\mathbf{f}_t^0\right\|\right]O(1) \\
&\le \max_{j=1,2}\left[\sup_\theta\left\|\mathbb{I}_{jt}(\theta)\mathbf{f}_t^0\right\|\right]\cdot\max_{j=1,2}\left[\sup_\theta\left\|\mathbb{I}_{jt}(\theta)\mathbf{f}_t^0\right\|\right]O(1):
\end{aligned}$$

the envelope function is $\mathcal{L}_{2\xi}$ bounded since by Schwarz's inequality and Assumption LT5(b)

$$\mathrm{E}\left|\max_{j=1,2}\left[\sup_\theta\left\|\mathbb{I}_{jt}(\theta)\mathbf{f}_t^0\right\|\right]\cdot\max_{j=1,2}\left[\sup_\theta\left\|\mathbb{I}_{jt}(\theta)\mathbf{f}_t^0\right\|\right]\right|^{2\xi} \le \left\{\mathrm{E}\left|\max_{j=1,2}\left[\sup_\theta\left\|\mathbb{I}_{jt}(\theta)\mathbf{f}_t^0\right\|\right]\right|^{4\xi}\right\}^{1/2}\left\{\mathrm{E}\left|\max_{j=1,2}\left[\sup_\theta\left\|\mathbb{I}_{jt}(\theta)\mathbf{f}_t^0\right\|\right]\right|^{4\xi}\right\}^{1/2} < \infty.$$

We then need to show that the log of the $\mathcal{L}_{2\xi}$ bracketing numbers $\mathbb{N}(\zeta)$ is integrable. For some $G<\infty$ and for all θ, there is some $\bar{\theta}$ such that $\left|\theta-\bar{\theta}\right|\le G\cdot\mathbb{N}(\zeta)^{-1}$. Set $\mathbb{N}(\zeta) = M^{1/\gamma}G\zeta^{-1/\gamma}$ and notice that

$$\begin{aligned}
\left[\mathrm{E}\left\|\mathbf{k}_{-,t}^{+0}(\theta) - \mathbf{k}_{-,t}^{+0}(\bar{\theta})\right\|^{2\xi}\right]^{1/(2\xi)} &= \left[\mathrm{E}\left\|\mathbf{f}_{-,t}^{+0}(\theta)f_{1t}^{+0} - \mathbf{f}_{-,t}^{+0}(\bar{\theta})f_{1t}^{+0}\right\|^{2\xi}\right]^{1/(2\xi)} \\
&= \left\{\mathrm{E}\left\|\left[\mathbb{I}_{1t}(\theta)\mathbf{f}_{-1,t}^{+0\prime}, \mathbb{I}_{2t}(\theta)\mathbf{f}_{-1,t}^{+0\prime}\right]'f_{1t}^{+0} - \left[\mathbb{I}_{1t}(\bar{\theta})\mathbf{f}_{-1,t}^{+0\prime}, \mathbb{I}_{2t}(\bar{\theta})\mathbf{f}_{-1,t}^{+0\prime}\right]'f_{1t}^{+0}\right\|^{2\xi}\right\}^{1/(2\xi)} \\
&= \left\{\mathrm{E}\left\|\begin{array}{c}\left[\mathbb{I}_{1t}(\theta)\left(\mathbf{H}_1^{+0}\mathbf{f}_t^0\right)', \mathbb{I}_{2t}(\theta)\left(\mathbf{H}_1^{+0}\mathbf{f}_t^0\right)'\right]'\mathbf{f}_t^{0\prime}\mathbf{h}_1^{+0} \\ -\left[\mathbb{I}_{1t}(\bar{\theta})\left(\mathbf{H}_1^{+0}\mathbf{f}_t^0\right)', \mathbb{I}_{2t}(\bar{\theta})\left(\mathbf{H}_1^{+0}\mathbf{f}_t^0\right)'\right]'\mathbf{f}_t^{0\prime}\mathbf{h}_1^{+0}\end{array}\right\|^{2\xi}\right\}^{1/(2\xi)} \\
&\le \left\{\mathrm{E}\left\|\left\{\left[\mathbb{I}_{1t}(\theta) - \mathbb{I}_{1t}(\bar{\theta})\right]\mathbf{f}_t^0\mathbf{f}_t^{0\prime}, \left[\mathbb{I}_{2t}(\theta) - \mathbb{I}_{2t}(\bar{\theta})\right]\mathbf{f}_t^0\mathbf{f}_t^{0\prime}\right\}'\right\|^{2\xi}\right\}^{1/(2\xi)}O(1) \\
&\le \left\{\mathrm{E}\left|\max_{j=1,2}\left\|\left[\mathbb{I}_{jt}(\theta) - \mathbb{I}_{jt}(\bar{\theta})\right]\mathbf{f}_t^0\mathbf{f}_t^{0\prime}\right\|\right|^{2\xi}\right\}^{1/(2\xi)}O(1).
\end{aligned}$$

By Assumption LT6,

$$\left\{\mathrm{E}\left|\max_{j=1,2}\left\|\left[\mathbb{I}_{jt}(\theta)-\mathbb{I}_{jt}(\bar{\theta})\right]\mathbf{f}_t^0\mathbf{f}_t^{0\prime}\right\|\right|^{2\xi}\right\}^{1/(2\xi)}\leq M\cdot\left|\theta-\bar{\theta}\right|^{\gamma}\leq M\cdot G^{\gamma}\cdot\mathbb{N}(\zeta)^{-\gamma}=\zeta,$$

so that $\mathbb{N}(\zeta)$ satisfies the definition of bracketing numbers: the log of $\mathbb{N}(\zeta)$ may be shown to be integrable as in the proof of Theorem 1 in Hansen (1996). It follows that $T^{-1/2}\sum_{t=1}^{T}\mathbf{f}_{-,t}^{+0}(\theta)f_{1t}^{+0}$ is stochastically equicontinuous and then $\hat{\mathbf{k}}_{-}(\theta)\Rightarrow\mathbf{k}_{-}^{0}(\theta)$. Let $\mathbf{0}_{(R^0-1)\times R^0}$ be the $\left(R^0-1\right)\times R^0$ zero matrix. Notice that

$$\begin{aligned}
\hat{\mathbf{M}}_{-}(\theta_1,\theta_2) &= T^{-1}\sum_{t=1}^{T}\tilde{\mathbf{f}}_{-,t}(\theta_1)\tilde{\mathbf{f}}_{-,t}(\theta_2)' \\
&= T^{-1}\sum_{t=1}^{T}\begin{bmatrix}\mathbb{I}_{1t}(\theta_1)\tilde{\mathbf{f}}_{-1,t}\\ \mathbb{I}_{2t}(\theta_1)\tilde{\mathbf{f}}_{-1,t}\end{bmatrix}\begin{bmatrix}\mathbb{I}_{1t}(\theta_2)\tilde{\mathbf{f}}_{-1,t}\\ \mathbb{I}_{2t}(\theta_2)\tilde{\mathbf{f}}_{-1,t}\end{bmatrix}' \\
&= T^{-1}\sum_{t=1}^{T}\left\{\begin{array}{c}\mathbb{I}_{1t}(\theta_1)\left[\mathbf{H}_1^{+0}\mathbf{f}_t^0+o_p(1)\right]\\ \mathbb{I}_{2t}(\theta_1)\left[\mathbf{H}_1^{+0}\mathbf{f}_t^0+o_p(1)\right]\end{array}\right\}\left\{\begin{array}{c}\mathbb{I}_{1t}(\theta_2)\left[\mathbf{H}_1^{+0}\mathbf{f}_t^0+o_p(1)\right]\\ \mathbb{I}_{2t}(\theta_2)\left[\mathbf{H}_1^{+0}\mathbf{f}_t^0+o_p(1)\right]\end{array}\right\}' \\
&= \begin{bmatrix}\mathbf{H}_1^{+0} & \mathbf{0}_{(R^0-1)\times R^0}\\ \mathbf{0}_{(R^0-1)\times R^0} & \mathbf{H}_1^{+0}\end{bmatrix}\hat{\mathbf{M}}(\theta_1,\theta_2)\begin{bmatrix}\mathbf{H}_1^{+0} & \mathbf{0}_{(R^0-1)\times R^0}\\ \mathbf{0}_{(R^0-1)\times R^0} & \mathbf{H}_1^{+0}\end{bmatrix}+o_p(1) \\
&\xrightarrow{p} \mathbf{M}_{-}^{0}(\theta_1,\theta_2)
\end{aligned}$$

uniformly in (θ_1,θ_2) by Assumption LT7, where

$$\mathbf{M}_{-}^{0}(\theta_1,\theta_2)=\begin{bmatrix}\mathbf{H}_1^{+0} & \mathbf{0}_{(R^0-1)\times R^0}\\ \mathbf{0}_{(R^0-1)\times R^0} & \mathbf{H}_1^{+0}\end{bmatrix}\mathbf{M}^{0}(\theta_1,\theta_2)\begin{bmatrix}\mathbf{H}_1^{+0} & \mathbf{0}_{(R^0-1)\times R^0}\\ \mathbf{0}_{(R^0-1)\times R^0} & \mathbf{H}_1^{+0}\end{bmatrix}.$$

The proof of the theorem is completed following similar steps as in the proof of Theorem 1 in Hansen (1996). ■

Proof of Lemma A.10. The proof is similar to that of Lemma 10 in Chen *et al.* (2014) and omitted. ■

Proof of Lemma A.11. The proof follows from Lemma A.10 and Assumption LT4. ■

References

[1] Acharya, V. V., L. H. Pedersen, T. Philippon, and M. Richardson (2010). Measuring Systemic Risk. Working Paper, New York University.

[2] Adrian, T., and M. K. Brunnermeier (2016). CoVaR. *American Economic Review*, 106 (7), 1705 − 1741.

[3] Alessi, L., M. Barigozzi, and M. Capasso (2010). Improved Penalization for Determining the Number of Factors in Approximate Factor Models. *Statistics and Probability Letters*, 80 (23 − 24), 1806−1813.

[4] Ang, A., and A. Timmermann (2012). Regime Changes and Financial Markets. *Annual Review of Financial Economics*, 4, 313 − 337.

[5] Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, 71 (1), 135 − 171.

[6] Bai, J., and Y. Liao (2016). Efficient Estimation of Approximate Factor Models via Penalized Maximum Likelihood. *Journal of Econometrics*, 191 (1), 1 − 18.

[7] Bai, J., and S. Ng (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70 (1), 191 − 221.

[8] Bates, B. J., M. Plagborg-Møller, J. H. Stock, and M. W. Watson (2013). Consistent Factor Estimation in Dynamic Factor Models with Structural Instability. *Journal of Econometrics*, 177 (2), 289 − 304.

[9] Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring Economic Policy Uncertainty. *Quarterly Journal of Economics*, 131 (4), 1593 − 1636.

[10] Billio, M., M. Getmansky, A. W. Lo, and L. Pelizzon (2012). Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors. *Journal of Financial Economics*, 104 (3), 535 − 559.

[11] Breitung, J., and S. Eickmeier (2011). Testing for Structural Breaks in Dynamic Factor Models. *Journal of Econometrics*, 163 (1), 71 − 84.

[12] Chamberlain, G., and M. Rothschild (1983). Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets. *Econometrica*, 51 (5), 1281 − 1304.

[13] Chan, K. S. (1993). Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model. *The Annals of Statistics*, 21(1), 520 − 533.

[14] Chan, K. S., and R. S. Tsay (1998). Limiting Properties of the Least Squares Estimator of a Continuous Threshold Autoregressive Model. *Biometrika*, 85 (2), 413 − 426.

[15] Chen, L. (2015). Estimating the Common Break Date in Large Factor Models. *Economics Letters*, 131, 70 − 74.

[16] Chen, L., J. J. Dolado, and J. Gonzalo (2014). Detecting Big Structural Breaks in Large Factor Models. *Journal of Econometrics*, 180 (1), 30 − 48.

[17] Cheng, X., Z. Liao, and F. Schorfheide (2015). Shrinkage Estimation of High-Dimensional Factor Models with Structural Instabilities. *Review of Economic Studies*, Forthcoming.

[18] Connor, G., and R. A. Korajczyk (1986). Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis. *Journal of Financial Economics*, 15 (3), 373 − 394.

[19] Connor, G., and R. A. Korajczyk (1988). Risk and Return in an Equilibrium APT: Application of a New Test Methodology. *Journal of Financial Economics*, 21 (2), 255 − 289.

[20] Connor, G., and R. A. Korajczyk (1993). A Test for the Number of Factors in an Approximate Factor Model. *Journal of Finance*, 48 (4), 1263 − 1288.

[21] Davies, R. B. (1977). Hypothesis Testing when a Nuisance Parameter is Present only under the Alternative. *Biometrika*, 64 (2), 247 − 254.

[22] Davies, R. B. (1987). Hypothesis Testing when a Nuisance Parameter is Present only under the Alternative. *Biometrika*, 74 (1), 33 − 43.

[23] Diebold, F. X., and K. Yilmaz (2014). On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms. *Journal of Econometrics*, 182 (1), 119 − 134.

[24] Doukhan, P., P. Massart, and E. Rio (1995). Invariance Principles for Absolutely Regular Empirical Processes. *Annales de l.I. H. P., Section B*, 31 (2), 393 − 427.

[25] Engle, R., and B. Kelly (2012). Dynamic Equicorrelation. *Journal of Business and Economic Statistics*, 30 (2), 212 − 228.

[26] Fan, J., Y. Liao, and M. Mincheva (2013). Large Covariance Estimation by Thresholding Principal Components Complements. *Journal of the Royal Statistical Society. Series B (Methodological)*, 75 (4), 603 − 680.

[27] Fan, J., Y. Liao, and W. Wang (2016a). Projected Principal Component Analysis in Factor Models. *The Annals of Statistics*, 44 (1), 219 − 254.

[28] Fan, J., W. Wang, and Y. Zhong (2016b). An l_∞ Eigenvector Perturbation Bound and Its Application to Robust Covariance Estimation. arXiv:1603.03516.

[29] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The Generalized Dynamic Factor Model: Identification and Estimation. *Review of Economics and Statistics*, 82 (4), 540 − 554.

[30] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2004). The Generalized Dynamic Factor Model: Consistency and Rates. *Journal of Econometrics*, 119 (2), 231 − 255.

[31] Forni, M., M. Hallin, M. Lippi, and P. Zaffaroni (2015). Dynamic Factor Models with Infinite-Dimensional Factor Spaces: One-Sided Representations. *Journal of Econometrics*, 185 (2), 359−371.

[32] Forni, M., and M. Lippi (2001). The Generalized Dynamic Factor Model: Representation Theory. *Econometric Theory*, 17 (6), 1113 − 1141.

[33] Hallin, M., and R. Liška (2007). Determining the Number of Factors in the General Dynamic Factor Model. *Journal of the American Statistical Association*, 102 (48), 603 − 617.

[34] Han, X., and A. Inoue (2015). Tests for Parameter Instability in Dynamic Factor Models. *Econometric Theory*, 31 (5), 1117 − 1152.

[35] Hansen, B. E. (1996). Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis. *Econometrica*, 64 (2), 413 − 430.

[36] Hansen, B. E. (1999). Threshold Effects in Non-Dynamic Panels: Estimation, Testing and Inference. *Journal of Econometrics*, 93 (2), 345 − 368.

[37] Hansen, B. E. (2000). Sample Splitting and Threshold Estimation. *Econometrica*, 68 (3), 575 − 603.

[38] Hansen, B. E. (2011). Threshold Autoregression in Economics. *Statistics and Its Interface*, 4 (2), 123 − 127.

[39] Jurado, K., S. C. Ludvigson, and S. Ng (2015). Measuring Uncertainty. *American Economic Review*, 105 (3), 1177 − 1216.

[40] Kapetanios, G. (2000). Small Sample Properties of the Conditional Least Squares Estimator in SETAR Models. *Economics Letters*, 69 (3), 267 − 276.

[41] Ludvigson, S. C., and S. Ng (2007). The Empirical Risk-Return Relation: A Factor Analysis Approach. *Journal of Financial Economics*, 83 (1), 171 − 222.

[42] Massacci, D. (2014). Multivariate Regime Switching Model with Flexible Threshold Variable. Working Paper, EIEF.

[43] Ng, S., and J. H. Wright (2013). Facts and Challenges from the Great Recession for Forecasting and Macroeconomic Modeling. *Journal of Economic Literature*, 51 (4), 1120 − 1154.

[44] Newey, W. K., and K. D. West (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autororrelation Consistent Covariance Matrix. *Econometrica*, 55 (3), 703 − 708.

[45] Pearson, K. (1900). On the Correlation of Characters not Quantitatively Measurable. *Royal Society Philosophical Transactions*, Series A, 195, 1 − 47.

[46] Peligrad, M. (1982). Invariance Principles for Mixing Sequences of Random Variables. *The Annals of Probability*, 10 (4), 968 − 981.

[47] Ross, S. (1976). The Arbitrage Theory of Capital Pricing. *Journal of Economic Theory*, 13 (3), 341 − 360.

[48] Stock, J. H., and M. W. Watson (2002). Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association*, 97 (460), 1167 − 1179.

[49] Timmermann, A. (2008). Elusive Return Predictability. *International Journal of Forecasting*, 24 (1), 1 − 18.

[50] Tong, H., and K. S. Lim (1980). Threshold Autoregression, Limit Cycles and Cyclical Data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42 (3), 245 − 292.

[51] Tsay, R. S. (1989). Testing and Modelling Threshold Autoregressive Processes. *Journal of the American Statistical Association*, 84 (405), 231 − 240.

[52] Tsay, R. S. (1998). Testing and Modelling Multivariate Threshold Models. *Journal of the American Statistical Association*, 93 (443), 1188 − 1202.

[53] Yamamoto, Y., and S. Tanaka (2015). Testing for Factor Loading Structural Change under Common Breaks. *Journal of Econometrics*, 189 (1), 187 − 206.

Table 1: Bias and RMSE in the case of the Estimator for $\theta^0 = 2$

This table presents results for the estimator for $\theta^0 = 2$. The DGP is detailed in Section 6.1. CSI denotes time homoskedastic factors and idiosyncratic components, and cross-sectionally independent idiosyncratic components. CSD denotes time homoskedastic factors and idiosyncratic components, and cross-sectionally dependent idiosyncratic components. CSDH denotes time heteroskedastic factors and idiosyncratic components, and cross-sectionally dependent idiosyncratic components.

Panel A: $\alpha^0 = 0.60$

N		25						50						100					
$\delta_i^0 > 0$		0.25		1.00		1.75		0.25		1.00		1.75		0.25		1.00		1.75	
T	**DGP**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**
	CSI	-0.0392	0.8826	-0.0055	0.0937	-0.0005	0.0193	-0.0120	0.8937	0.0014	0.1016	-0.0010	0.0203	0.0297	0.8692	-0.0026	0.0996	-0.0013	0.0157
100	CSD	-0.0121	0.9323	-0.0048	0.2066	-0.0007	0.0363	0.0006	0.9168	-0.0021	0.1672	-0.0013	0.0265	0.0437	0.8987	-0.0058	0.1210	-0.0015	0.0199
	CSDH	-0.0323	0.9329	-0.0046	0.2385	-0.0007	0.0400	0.0007	0.9335	0.0013	0.1935	-0.0009	0.0295	0.0496	0.8995	-0.0075	0.1454	-0.0018	0.0234
	CSI	-0.0101	0.7709	-0.0004	0.0156	0.0000	0.0070	-0.0162	0.7795	-0.0013	0.0194	-0.0002	0.0067	-0.0095	0.7425	-0.0007	0.0113	-0.0002	0.0048
200	CSD	0.0131	0.8340	0.0003	0.0316	-0.0003	0.0113	-0.0189	0.8216	-0.0027	0.0304	-0.0005	0.0086	0.0114	0.7616	-0.0011	0.0179	-0.0003	0.0059
	CSDH	0.0165	0.8455	0.0006	0.0501	-0.0002	0.0128	-0.0040	0.8341	-0.0030	0.0382	-0.0005	0.0092	-0.0062	0.7781	-0.0013	0.0187	-0.0004	0.0060
	CSI	-0.0029	0.4862	0.0000	0.0047	-0.0001	0.0028	0.0050	0.5616	0.0000	0.0055	-0.0002	0.0028	-0.0277	0.4889	0.0000	0.0030	0.0000	0.0030
400	CSD	0.0138	0.6241	0.0002	0.0085	-0.0001	0.0036	-0.0096	0.6405	-0.0001	0.0079	-0.0002	0.0037	-0.0346	0.5612	0.0000	0.0054	-0.0001	0.0035
	CSDH	-0.0044	0.6503	0.0004	0.0094	-0.0001	0.0037	0.0130	0.6523	-0.0002	0.0083	-0.0001	0.0034	-0.0322	0.5659	-0.0001	0.0055	-0.0001	0.0038

Panel B: $\alpha^0 = 1.00$

N		25						50						100					
$\delta_i^0 > 0$		0.25		1.00		1.75		0.25		1.00		1.75		0.25		1.00		1.75	
T	**DGP**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**
	CSI	-0.0412	0.7581	-0.0046	0.0398	-0.0032	0.0254	-0.0100	0.6852	-0.0020	0.0269	-0.0015	0.0169	0.0090	0.6311	-0.0021	0.0229	-0.0012	0.0119
100	CSD	-0.0577	0.8373	-0.0169	0.1347	-0.0071	0.0424	-0.0208	0.7609	-0.0033	0.0389	-0.0019	0.0187	0.0208	0.6676	-0.0033	0.0303	-0.0019	0.0154
	CSDH	-0.0643	0.8585	-0.0145	0.1694	-0.0090	0.0477	-0.0260	0.7686	-0.0049	0.0569	-0.0021	0.0191	0.0205	0.6913	-0.0041	0.0330	-0.0018	0.0163
	CSI	-0.0223	0.4862	-0.0013	0.0122	-0.0006	0.0091	-0.0240	0.3901	-0.0006	0.0077	-0.0003	0.0041	-0.0212	0.2801	-0.0005	0.0066	-0.0002	0.0039
200	CSD	-0.0118	0.6344	-0.0041	0.0245	-0.0023	0.0145	-0.0292	0.4902	-0.0009	0.0110	-0.0005	0.0061	-0.0198	0.3490	-0.0006	0.0080	-0.0004	0.0052
	CSDH	-0.0216	0.6536	-0.0044	0.0261	-0.0023	0.0146	-0.0396	0.4931	-0.0012	0.0114	-0.0005	0.0065	-0.0274	0.3820	-0.0006	0.0082	-0.0004	0.0050
	CSI	-0.0106	0.1988	-0.0004	0.0042	-0.0002	0.0027	-0.0011	0.0990	-0.0001	0.0031	0.0000	0.0020	-0.0034	0.0397	-0.0001	0.0019	0.0000	0.0011
400	CSD	-0.0075	0.3536	-0.0010	0.0074	-0.0003	0.0036	-0.0048	0.1860	-0.0002	0.0036	-0.0001	0.0025	-0.0041	0.0542	-0.0001	0.0025	0.0000	0.0010
	CSDH	-0.0070	0.3734	-0.0008	0.0072	-0.0003	0.0040	-0.0059	0.1982	-0.0002	0.0033	-0.0001	0.0024	-0.0051	0.0726	-0.0001	0.0027	-0.0001	0.0014

Table 2: MSE in the case of the Estimator for c_{it}^{0s}

This table presents results for the estimator for

$$c_{it}^{0s} = \mathbb{I}\left(z_t^s \le \theta^0\right)\lambda_{1i}^0 f_t^{0s} + \mathbb{I}\left(z_t^s > \theta^0\right)\lambda_{2i}^0 f_t^{0s}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T.$$

The DGP is the same as in Table 1. CSI, CSD and CSDH denote the same scenarios as in Table 1.

Panel A: $\alpha^0 = 0.60$

		Unfeasible Estimator									**Feasible Estimator**								
	N	25			50			100			25			50			100		
	$\delta_i^0 > 0$	0.25	1.00	1.75	0.25	1.00	1.75	0.25	1.00	1.75	0.25	1.00	1.75	0.25	1.00	1.75	0.25	1.00	1.75
T	**DGP**																		
100	CSI	0.0595	0.0570	0.0549	0.0513	0.0515	0.0515	0.0367	0.0369	0.0370	0.0705	0.0585	0.0551	0.0609	0.0530	0.0516	0.0426	0.0375	0.0371
	CSD	0.1114	0.1052	0.1000	0.0623	0.0626	0.0626	0.0415	0.0421	0.0426	0.1273	0.1106	0.1009	0.0746	0.0659	0.0629	0.0480	0.0431	0.0427
	CSDH	0.1125	0.1061	0.1007	0.0625	0.0627	0.0627	0.0415	0.0421	0.0426	0.1292	0.1120	0.1019	0.0753	0.0665	0.0631	0.0482	0.0433	0.0427
200	CSI	0.0467	0.0445	0.0424	0.0367	0.0370	0.0371	0.0222	0.0224	0.0226	0.0523	0.0446	0.0425	0.0416	0.0372	0.0371	0.0251	0.0224	0.0226
	CSD	0.0941	0.0882	0.0831	0.0468	0.0472	0.0474	0.0267	0.0273	0.0279	0.1023	0.0888	0.0833	0.0530	0.0477	0.0475	0.0299	0.0274	0.0279
	CSDH	0.0942	0.0882	0.0831	0.0468	0.0472	0.0474	0.0266	0.0273	0.0278	0.1028	0.0889	0.0833	0.0532	0.0477	0.0475	0.0300	0.0274	0.0278
400	CSI	0.0406	0.0384	0.0365	0.0292	0.0295	0.0297	0.0149	0.0151	0.0153	0.0428	0.0385	0.0365	0.0313	0.0296	0.0297	0.0160	0.0153	0.0153
	CSD	0.0863	0.0805	0.0756	0.0388	0.0393	0.0395	0.0192	0.0200	0.0205	0.0902	0.0806	0.0757	0.0418	0.0393	0.0395	0.0207	0.0200	0.0205
	CSDH	0.0860	0.0802	0.0754	0.0392	0.0392	0.0395	0.0192	0.0199	0.0205	0.0900	0.0803	0.0754	0.0418	0.0393	0.0395	0.0206	0.0199	0.0205

Panel B: $\alpha^0 = 1.00$

		Unfeasible Estimator									**Feasible Estimator**								
	N	25			50			100			25			50			100		
	$\delta_i^0 > 0$	0.25	1.00	1.75	0.25	1.00	1.75	0.25	1.00	1.75	0.25	1.00	1.75	0.25	1.00	1.75	0.25	1.00	1.75
T	**DGP**																		
100	CSI	0.0594	0.0581	0.0574	0.0509	0.0502	0.0498	0.0366	0.0366	0.0367	0.0701	0.0586	0.0576	0.0605	0.0504	0.0499	0.0422	0.0367	0.0367
	CSD	0.1117	0.1094	0.1086	0.0625	0.0631	0.0632	0.0416	0.0423	0.0427	0.1273	0.1125	0.1094	0.0754	0.0637	0.0485	0.0480	0.0425	0.0428
	CSDH	0.1125	0.1099	0.1090	0.0627	0.0631	0.0632	0.0416	0.0422	0.0426	0.1290	0.1137	0.1100	0.0759	0.0639	0.0634	0.0482	0.0424	0.0427
200	CSI	0.0469	0.0460	0.0455	0.0364	0.0359	0.0356	0.0221	0.0222	0.0224	0.0511	0.0460	0.0455	0.0394	0.0359	0.0356	0.0235	0.0223	0.0224
	CSD	0.0949	0.0934	0.0928	0.0472	0.0481	0.0484	0.0268	0.0277	0.0281	0.1020	0.0938	0.0930	0.0520	0.0482	0.0407	0.0287	0.0277	0.0281
	CSDH	0.0949	0.0932	0.0927	0.0472	0.0481	0.0484	0.0268	0.0276	0.0281	0.1024	0.0937	0.0928	0.0522	0.0482	0.0484	0.0288	0.0276	0.0281
400	CSI	0.0409	0.0402	0.0398	0.0289	0.0285	0.0282	0.0148	0.0150	0.0152	0.0419	0.0402	0.0398	0.0293	0.0285	0.0282	0.0150	0.0150	0.0152
	CSD	0.0873	0.0861	0.0857	0.0393	0.0404	0.0407	0.0194	0.0203	0.0208	0.0898	0.0862	0.0857	0.0403	0.0404	0.0407	0.0196	0.0203	0.0208
	CSDH	0.0869	0.0857	0.0853	0.0392	0.0402	0.0406	0.0194	0.0203	0.0207	0.0896	0.0858	0.0853	0.0403	0.0403	0.0406	0.0196	0.0203	0.0208

Table 3: Model Selection Criteria, $R^0 = 2$, $\alpha^0 = 0.60$

This table presents results for the model selection criteria in (13). The DGP is detailed in Section 6.2. CSI, CSD and CSDH denote the same scenarios as in Table 1.

Panel A: $IC_{p1}(R, R)$

		Unfeasible Estimator									Feasible Estimator								
N		25			50			100			25			50			100		
$\delta_i^0 > 0$		0.25	1.00	1.75	0.25	1.00	1.75	0.25	1.00	1.75	0.25	1.00	1.75	0.25	1.00	1.75	0.25	1.00	1.75
T	**DGP**																		
100	CSI	2.7835	2.7995	2.8070	2.9285	2.9230	2.9175	2.0145	2.0145	2.0145	2.7955	2.7935	2.8125	2.9750	2.9325	2.9280	2.0175	2.0155	2.0270
	CSD	3.0600	3.0885	3.1185	4.1330	4.1230	4.1135	2.0640	2.0645	2.0640	3.0670	3.0805	3.1440	4.1735	4.1355	4.1435	2.0750	2.0695	2.0815
	CSDH	3.0415	3.0845	3.1090	4.1385	4.1260	4.1235	2.1155	2.1160	2.1160	3.0560	3.0750	3.1365	4.1955	4.1415	4.1635	2.1230	2.1205	2.1300
200	CSI	2.9075	2.9195	2.9235	2.9755	2.9630	2.9600	2.0130	2.0135	2.0135	2.9060	2.9185	2.9235	2.9785	2.9655	2.9625	2.0140	2.0135	2.0135
	CSD	3.0465	3.0610	3.0700	4.4945	4.4695	4.4580	2.0585	2.0590	2.0590	3.0450	3.0560	3.0715	4.5345	4.4690	4.4610	2.0650	2.0610	2.0590
	CSDH	3.0325	3.0575	3.0710	4.4460	4.4230	4.4065	2.1040	2.1020	2.1020	3.0285	3.0545	3.0725	4.4875	4.4265	4.4105	2.1095	2.1040	2.1030
400	CSI	2.9655	2.9710	2.9765	2.9960	2.9820	2.9770	2.0040	2.0040	2.0040	2.9645	2.9710	2.9765	2.9980	2.9815	2.9915	2.0045	2.0040	2.0040
	CSD	3.0085	3.0145	3.0160	5.1125	5.1185	5.1200	2.0360	2.0360	2.0355	3.0100	3.0145	3.0160	5.1195	5.1185	5.1200	2.0370	2.0360	2.0355
	CSDH	3.0100	3.0190	3.0260	4.9345	4.9235	4.9150	2.0770	2.0780	2.0780	3.0095	3.0190	3.0260	4.9475	4.9260	4.9155	2.0765	2.0780	2.0780

Panel B: $IC_{p2}(R, R)$

		Unfeasible Estimator									Feasible Estimator								
N		25			50			100			25			50			100		
$\delta_i^0 > 0$		0.25	1.00	1.75	0.25	1.00	1.75	0.25	1.00	1.75	0.25	1.00	1.75	0.25	1.00	1.75	0.25	1.00	1.75
T	**DGP**																		
100	CSI	2.6035	2.6170	2.6295	2.3210	2.3215	2.3215	2.0000	2.0000	2.0000	2.6090	2.6150	2.6340	2.3130	2.3160	2.3240	2.0000	2.0000	2.0010
	CSD	2.8160	2.8585	2.8870	3.3290	3.3220	3.3115	2.0010	2.0005	2.0010	2.8265	2.8405	2.8985	3.3440	3.3240	3.3310	2.0010	2.0005	2.0020
	CSDH	2.7915	2.8365	2.8620	3.3860	3.3715	3.3640	2.0050	2.0050	2.0050	2.7915	2.8270	2.8795	3.3975	3.3645	3.3800	2.0050	2.0060	2.0070
200	CSI	2.8290	2.8375	2.8475	2.5550	2.5510	2.5505	2.0005	2.0005	2.0005	2.8235	2.8370	2.8475	2.5585	2.5530	2.5515	2.0005	2.0005	2.0005
	CSD	2.9700	2.9965	3.0085	3.7905	3.7625	3.7460	2.0045	2.0045	2.0040	2.9685	2.9940	3.0095	3.7925	3.7700	3.7525	2.0045	2.0050	2.0045
	CSDH	2.9395	2.9710	2.9915	3.8425	3.8190	3.8005	2.0205	2.0205	2.0205	2.9430	2.9670	2.9910	3.8400	3.8190	3.8025	2.0195	2.0205	2.0200
400	CSI	2.9450	2.9530	2.9560	2.7385	2.7375	2.7385	2.0000	2.0000	2.0000	2.9440	2.9530	2.9560	2.7405	2.7375	2.7435	2.0000	2.0000	2.0000
	CSD	3.0005	3.0045	3.0060	4.5885	4.5610	4.5355	2.0045	2.0045	2.0045	3.0015	3.0045	3.0060	4.5850	4.5625	4.5355	2.0055	2.0045	2.0045
	CSDH	2.9905	3.0025	3.0090	4.4880	4.4625	4.4420	2.0260	2.0260	2.0260	2.9915	3.0025	3.0090	4.4980	4.4630	4.4410	2.0255	2.0260	2.0260

Panel C: $IC_{p3}(R, R)$

		Unfeasible Estimator									Feasible Estimator								
N		25			50			100			25			50			100		
$\delta_i^0 > 0$		0.25	1.00	1.75	0.25	1.00	1.75	0.25	1.00	1.75	0.25	1.00	1.75	0.25	1.00	1.75	0.25	1.00	1.75
T	**DGP**																		
100	CSI	3.0840	3.1030	3.1145	5.8240	5.8175	5.8170	7.8165	7.8160	7.8105	3.0845	3.1035	3.1325	5.9020	5.8490	5.8395	7.9240	7.8685	7.8510
	CSD	3.8870	3.9405	3.9795	6.1995	6.1935	6.1930	7.9975	7.9980	7.9980	3.9325	3.9690	4.0845	6.3015	6.2755	6.2540	7.9995	7.9990	7.9990
	CSDH	3.9385	3.9845	4.0190	6.2195	6.2110	6.2095	7.9980	7.9980	7.9980	3.9740	4.0115	4.1110	6.3425	6.3010	6.2840	7.9990	8.0000	8.0000
200	CSI	2.9840	2.9865	2.9915	5.1095	5.1020	5.0905	3.2865	3.2795	3.2790	2.9830	2.9865	2.9915	5.1365	5.1050	5.0920	3.3305	3.2840	3.2815
	CSD	3.2645	3.2860	3.2975	5.8315	5.8385	5.8405	4.8485	4.8235	4.8080	3.2685	3.2845	3.3040	5.8510	5.8400	5.8420	4.9595	4.8395	4.8090
	CSDH	3.2925	3.3195	3.3290	5.7225	5.7300	5.7360	4.8625	4.8385	4.8245	3.2940	3.3165	3.3380	5.7460	5.7335	5.7395	4.9860	4.8465	4.8255
400	CSI	2.9945	2.9975	2.9980	4.5025	4.4815	4.4600	2.4670	2.4660	2.4660	2.9945	2.9975	2.9980	4.5170	4.4810	4.4845	2.4715	2.4650	2.4660
	CSD	3.0495	3.0585	3.0650	5.8500	5.8595	5.8700	2.9135	2.9130	2.9100	3.0460	3.0585	3.0640	5.8510	5.8590	5.8695	2.9310	2.9145	2.9100
	CSDH	3.0660	3.0820	3.0875	5.7320	5.7480	5.7535	2.9285	2.9230	2.9190	3.0660	3.0805	3.0875	5.7370	5.7465	5.7525	2.9355	2.9220	2.9175

Table 4: Linearity Test, $\alpha^0 = 0.60$

This table presents results for the linearity test proposed in Section 5. The DGP is detailed in Section 6.3. CSI, CSD and CSDH denote the same scenarios as in Table 1.

Panel A: 5% Level

N		25					50					100				
		Size, $R^0=2$		**Power,** $R^0=1$			**Size,** $R^0=2$		**Power,** $R^0=1$			**Size,** $R^0=2$		**Power,** $R^0=1$		
Statistic		$\sup\widehat{LM}^{\text{HC}}$	$\sup\widehat{LM}^{\text{HAC}}$	$\sup\widehat{LM}^{\text{HAC}}$			$\sup\widehat{LM}^{\text{HC}}$	$\sup\widehat{LM}^{\text{HAC}}$	$\sup\widehat{LM}^{\text{HAC}}$			$\sup\widehat{LM}^{\text{HC}}$	$\sup\widehat{LM}^{\text{HAC}}$	$\sup\widehat{LM}^{\text{HAC}}$		
ρ_f		0.00	0.50	0.50			0.00	0.50	0.50			0.00	0.50	0.50		
$\delta_i^0 > 0$		-	-	0.25	1.00	1.75	-	-	0.25	1.00	1.75	-	-	0.25	1.00	1.75
T	**DGP**															
100	CSI	0.0205	0.0055	0.0090	0.0190	0.0875	0.0215	0.0060	0.0090	0.0645	0.4985	0.0255	0.0100	0.0150	0.0970	0.6545
	CSD	0.0245	0.0085	0.0095	0.0265	0.1660	0.0210	0.0050	0.0090	0.0655	0.4035	0.0275	0.0095	0.0115	0.0830	0.5085
	CSDH	0.0315	0.0095	0.0100	0.0285	0.1610	0.0360	0.0060	0.0080	0.0600	0.3925	0.0340	0.0105	0.0115	0.0760	0.4580
200	CSI	0.0295	0.0245	0.0290	0.0625	0.1745	0.0320	0.0200	0.0325	0.1500	0.7330	0.0365	0.0245	0.0385	0.3555	0.9460
	CSD	0.0280	0.0240	0.0280	0.0775	0.2840	0.0345	0.0175	0.0285	0.1305	0.6045	0.0365	0.0240	0.0455	0.3740	0.9170
	CSDH	0.0440	0.0285	0.0220	0.0755	0.2850	0.0455	0.0205	0.0275	0.1300	0.6030	0.0440	0.0255	0.0420	0.3485	0.8860
400	CSI	0.0335	0.0485	0.0540	0.0970	0.2485	0.0385	0.0485	0.0510	0.2375	0.8520	0.0435	0.0405	0.0830	0.6720	0.9985
	CSD	0.0370	0.0495	0.0595	0.1390	0.4160	0.0385	0.0485	0.0655	0.1825	0.7245	0.0435	0.0405	0.0900	0.7455	0.9975
	CSDH	0.0520	0.0495	0.0560	0.1320	0.4190	0.0570	0.0470	0.0635	0.1800	0.7300	0.0570	0.0415	0.0820	0.7375	0.9935

Panel B: 10% Level

N		25					50					100				
		Size, $R^0=2$		**Power,** $R^0=1$			**Size,** $R^0=2$		**Power,** $R^0=1$			**Size,** $R^0=2$		**Power,** $R^0=1$		
Statistic		$\sup\widehat{LM}^{\text{HC}}$	$\sup\widehat{LM}^{\text{HAC}}$	$\sup\widehat{LM}^{\text{HAC}}$			$\sup\widehat{LM}^{\text{HC}}$	$\sup\widehat{LM}^{\text{HAC}}$	$\sup\widehat{LM}^{\text{HAC}}$			$\sup\widehat{LM}^{\text{HC}}$	$\sup\widehat{LM}^{\text{HAC}}$	$\sup\widehat{LM}^{\text{HAC}}$		
ρ_f		0.00	0.50	0.50			0.00	0.50	0.50			0.00	0.50	0.50		
$\delta_i^0 > 0$		-	-	0.25	1.00	1.75	-	-	0.25	1.00	1.75	-	-	0.25	1.00	1.75
T	**DGP**															
100	CSI	0.0630	0.0255	0.0325	0.0515	0.1635	0.0595	0.0240	0.0375	0.1340	0.5980	0.0585	0.0260	0.0510	0.1885	0.7365
	CSD	0.0610	0.0270	0.0370	0.0735	0.2510	0.0600	0.0230	0.0405	0.1295	0.5015	0.0565	0.0250	0.0465	0.1735	0.6195
	CSDH	0.0735	0.0300	0.0325	0.0680	0.2525	0.0785	0.0245	0.0355	0.1245	0.4840	0.0695	0.0295	0.0420	0.1660	0.5840
200	CSI	0.0690	0.0635	0.0765	0.1240	0.2475	0.0755	0.0535	0.0750	0.2245	0.7795	0.0725	0.0585	0.0940	0.4615	0.9620
	CSD	0.0720	0.0670	0.0695	0.1465	0.3775	0.0765	0.0530	0.0810	0.2060	0.6640	0.0740	0.0590	0.1070	0.4845	0.9395
	CSDH	0.0990	0.0665	0.0710	0.1465	0.3700	0.0945	0.0570	0.0715	0.2035	0.6655	0.0920	0.0630	0.0880	0.4620	0.9150
400	CSI	0.0885	0.0980	0.1170	0.1775	0.3545	0.0920	0.0975	0.1135	0.3270	0.8750	0.0895	0.0860	0.1510	0.7335	0.9985
	CSD	0.0880	0.0970	0.1145	0.2210	0.4945	0.0920	0.0950	0.1220	0.2645	0.7680	0.0915	0.0870	0.1650	0.8075	0.9975
	CSDH	0.1055	0.1105	0.1130	0.2190	0.5025	0.1060	0.1070	0.1120	0.2730	0.7715	0.1145	0.0915	0.1580	0.8035	0.9955

Table 5: Empirical Application, Estimation Results, 1985 - 2014

This table presents results from the empirical application of the model in (1). The vector $\mathbf{x}_t$ is made of the 147 updated monthly financial variables employed in Jurado *et al.* (2015). The threshold variable z_t is the lagged index of economic policy uncertainty proposed in Baker *et al.* (2016). The model is estimated over the period $1985:01-2014:12$, a total of 360 observations. $\hat{\theta}$ is the point estimate of the threshold parameter θ^0 and $\hat{\pi}=T^{-1}\sum_{t=1}^{T}\mathbb{I}\left(z_t\leq\hat{\theta}\right)$. The optimal number of factors $\hat{R}$ is estimated according to the selection criteria $IC_{p1}(R,R)$, $IC_{p2}(R,R)$ and $IC_{p3}(R,R)$ in (13). The connectedness measures $C_1\left(\hat{R}\right)$ and $C_2\left(\hat{R}\right)$ are as in (20).

$\hat{\theta}$	131.413		
$\hat{\pi}$	0.783		
$1-\hat{\pi}$	0.217		
	$IC_{p1}(R,R)$	$IC_{p2}(R,R)$	$IC_{p3}(R,R)$
$\hat{R}$	3	3	6
$C_1\left(\hat{R}\right)$	0.678	0.678	0.736
$C_2\left(\hat{R}\right)$	0.865	0.865	0.898

Figure 1: Empirical Application, High Economic Policy Uncertainty Regime, 1985 - 2014

This figure shows the high economic policy uncertainty regime, as identified by the sequence $\left\{\mathbb{I}\left(z_t>\hat{\theta}\right)=1\right\}_{t=1}^{T}$, where $\hat{\theta}=131.413$ is the point estimate of the threshold parameter θ^0.